CORYNEBACTERIUM GLUTAMICUM GENES ENCODING METABOLIC PATHWAY PROTEINS

Related Applications

The present application is an continuation in part of U.S. Patent Application 09/606,740, filed June 23, 2000. This application is also a continuation in part of U.S. Patent Application 09/603,124, filed June 23, 2000. The present application claims priority to prior filed U.S. Provisional Patent Application Serial No. 60/141031, filed June 25 , 1999, U.S. Provisional Patent Application Serial No. 60/142101, filed July 2, 1999, U.S. Provisional Patent Application Serial No. 60/148613, filed August 12, 1999, U.S. Provisional Patent Application Serial No. 60/187970, filed March 9, 2000, and also to German Patent Application No. 19931420.9, filed July 8, 1999. The entire contents of all of the aforementioned applications are hereby expressly incorporated herein by this reference.

Background of the Invention

Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through large-scale culture of bacteria developed to produce and secrete large quantities of a particular desired molecule. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

Summary of the Invention

The invention provides novel bacterial nucleic acid molecules which have a variety of uses. These uses include the identification of microorganisms which can be

used to produce fine chemicals (*e.g.*, amino acids, such as, for example, lysine and methionine), the modulation of fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or related bacteria, as reference points for mapping the *C. glutamicum* genome, and as markers for transformation. These novel nucleic acid molecules encode proteins, referred to herein as metabolic pathway (MP) proteins.

C. glutamicum is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The MP nucleic acid molecules of the invention, therefore, can be used to identify microorganisms which can be used to produce fine chemicals, *e.g.*, by fermentation processes. Modulation of the expression of the MP nucleic acids of the invention, or modification of the sequence of the MP nucleic acid molecules of the invention, can be used to modulate the production of one or more fine chemicals from a microorganism (*e.g.*, to improve the yield or production of one or more fine chemicals from a Corynebacterium or Brevibacterium species). In a preferred embodiment, the MP genes of the invention are combined with one or more genes involved in the same or different metabolic pathway to modulate the production of one or more fine chemicals from a microorganism.

The MP nucleic acids of the invention may also be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof, or to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to species pathogenic in humans, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria); the detection of such organisms is of significant clinical relevance.

The MP nucleic acid molecules of the invention may also serve as reference points for mapping of the *C. glutamicum* genome, or of genomes of related organisms.

Similarly, these molecules, or variants or portions thereof, may serve as markers for genetically engineered Corynebacterium or Brevibacterium species.

The MP proteins encoded by the novel nucleic acid molecules of the invention are capable of, for example, performing an enzymatic step involved in the metabolism of certain fine chemicals, including amino acids, *e.g.*, lysine and methionine, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey *et al.*, U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (*e.g.*, *lactofermentum*) (Yoshihama et al, *J. Bacteriol.* 162: 591-597 (1985); Katsumata *et al.*, *J. Bacteriol.* 159: 306-311 (1984); and Santamaria *et al.*, *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals.

This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation. Specifically, alterations in *C. glutamicum* metabolic pathways for amino acids, *e.g.*, lysine and methionine, vitamins, cofactors, nucleotides, and trehalose may have a direct impact on the overall production of one or more of these desired compounds from this organism. For example, optimizing the activity of a lysine or a methionine biosynthetic pathway protein or decreasing the activity of a lysine or methionine degradative pathway protein may result in an increase in the yield or efficiency of production of lysine or methionine from such an engineered organism. Alterations in the proteins involved in these metabolic pathways may also have an indirect impact on the production or efficiency of production of a desired fine chemical. For example, a reaction which is in competition for an intermediate necessary for the production of a desired molecule may be eliminated, or a pathway necessary for the production of a particular intermediate for a desired compound may be optimized. Further, modulations in the biosynthesis or degradation of, for example, an amino acid, *e.g.*, lysine or methionine, a vitamin, or a nucleotide may increase the overall ability of the microorganism to rapidly grow and divide, thus increasing the number and/or production capacities of the microorganism in culture and thereby increasing the possible yield of the desired fine chemical.

The nucleic acid and protein molecules of the invention, alone or in combination with one or more nucleic acid and protein molecules of the same or different metabolic pathway, may be utilized to directly improve the production or efficiency of production of one or more desired fine chemicals from *Corynebacterium glutamicum* (*e.g.*, methionine or lysine). Using recombinant genetic techniques well known in the art, one or more of the biosynthetic or degradative enzymes of the invention for amino acids, *e.g.*, lysine and methionine, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose may be manipulated such that its function is modulated. For example, a biosynthetic enzyme may be improved in efficiency, or its allosteric control region destroyed such that feedback inhibition of production of the compound is prevented. Similarly, a degradative enzyme may be deleted or modified by substitution, deletion, or addition such that its degradative activity is lessened for the desired compound without impairing the viability of the cell. In each case, the overall yield or rate of production of the desired fine chemical may be increased.

It is also possible that such alterations in the protein and nucleotide molecules of the invention may improve the production of other fine chemicals besides the amino acids, *e.g.*, lysine and methionine, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose through indirect mechanisms. Metabolism of any one compound is necessarily intertwined with other biosynthetic and degradative pathways within the cell, and necessary cofactors, intermediates, or substrates in one pathway are likely supplied or limited by another such pathway. Therefore, by modulating the activity of one or more of the proteins of the invention, the production or efficiency of activity of another fine chemical biosynthetic or degradative pathway may be impacted. For example, amino acids serve as the structural units of all proteins, yet may be present intracellularly in levels which are limiting for protein synthesis; therefore, by increasing the efficiency of production or the yields of one or more amino acids within the cell, proteins, such as biosynthetic or degradative proteins, may be more readily synthesized. Likewise, an alteration in a metabolic pathway enzyme such that a particular side reaction becomes more or less favored may result in the over- or under-production of one or more compounds which are utilized as intermediates or substrates for the production of a desired fine chemical.

This invention provides novel nucleic acid molecules which encode proteins, referred to herein as metabolic pathway ("MP") proteins, which are capable of, for

example, performing an enzymatic step involved in the metabolism of molecules important for the normal functioning of cells, such as amino acids, *e.g.*, lysine and methionine, vitamins, cofactors, nucleotides and nucleosides, or trehalose. Nucleic acid molecules encoding an MP protein are referred to herein as MP nucleic acid molecules. In a preferred embodiment, an MP protein, alone or in combination with one or more proteins of the same or different metabolic pathway, performs an enzymatic step related to the metabolism of one or more of the following: amino acids, *e.g.*, lysine and methionine, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose. Examples of such proteins include those encoded by the genes set forth in Table 1.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (*e.g.*, cDNAs, DNAs, or RNAs) comprising a nucleotide sequence encoding an MP protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of MP-encoding nucleic acid (*e.g.*, DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth as the odd-numbered SEQ ID NO in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5), or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to a nucleotide sequence set forth as an odd-numbered SEQ ID NO in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5), or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth as an even-numbered SEQ ID NO in the Sequence Listing (*e.g.*, SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6). The preferred MP proteins of the present invention also preferably possess at least one of the MP activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence

which is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence having an even-numbered SEQ ID NO in the Sequence Listing, such as SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6), *e.g.*, sufficiently homologous to an amino acid sequence of the invention such that the protein or portion thereof maintains an MP activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to perform an enzymatic reaction in a amino acid, *e.g.*, lysine or methionine, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to an amino acid sequence of the invention (*e.g.*, an entire amino acid sequence selected from those having an even-numbered SEQ ID NO in the Sequence Listing, such as SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of the invention (encoded by an open reading frame shown in the corresponding odd-numbered SEQ ID NO in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (*e.g.*, an MP fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of the invention (*e.g.*, a sequence of one of the even-numbered SEQ ID NOs in the Sequence Listing, such as SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6) and is able to catalyze a reaction in a metabolic pathway for an amino acid, *e.g.*, lysine or methionine, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose, or one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of the invention (*e.g.*, a sequence of an odd-

numbered SEQ ID NO in the Sequence Listing, such as SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5). Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* MP protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, *e.g.*, recombinant expression vectors, containing the nucleic acid molecules of the invention, alone or in combination with one or more nucleic acid molecules involved in the same or different pathway, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an MP protein by culturing the host cell in a suitable medium. The MP protein can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which one or more MP genes, alone or in combination with one or more genes involved in the same or different metabolic pathway, have been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding one or more wild-type or mutated MP sequences as transgenes alone or in combination with one or more nucleic acid molecules involved in the same or different metabolic pathway. In another embodiment, one or more endogenous MP genes within the genome of the microorganism have been altered, *e.g.*, functionally disrupted, by homologous recombination with one or more altered MP genes. In another embodiment, one or more endogenous or introduced MP genes, alone or in combination with one or more genes of the same or different metabolic pathway in a microorganism have been altered by one or more point mutations, deletions, or inversions, but still encode functional MP proteins. In still another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or inducer) of one or more MP genes in a microorganism, alone or in combination with one or more MP genes or in combination with one or more genes of the same or different metabolic pathway, has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of one or more MP genes is modulated. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine and

methionine being particularly preferred. In a particularly preferred embodiment, the MP gene is the *metZ* gene (SEQ ID NO:1), *metC* gene (SEQ ID NO:3), or the RXA00657 gene (SEQ ID NO:5), alone or in combination with one or more MP genes of the invention or in combination with one or more genes involved in methionine and/or lysine metabolism.

In another aspect, the invention provides a method of identifying the presence or activity of *Cornyebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (*e.g.*, the sequences set forth in Table 1 and in the Sequence Listing as SEQ ID NOs 1 through 122) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject.

Still another aspect of the invention pertains to an isolated MP protein or portion, *e.g.*, biologically active portion, thereof. In a preferred embodiment, the isolated MP protein or portion thereof, alone or in combination with one or more MP proteins of the invention or in combination with one or more proteins of the same or different metabolic pathway, can catalyze an enzymatic reaction involved in one or more pathways for the metabolism of an amino acid, *e.g.*, lysine or methionine, a vitamin, a cofactor, a nutraceutical, a nucleotide, a nucleoside, or trehalose. In another preferred embodiment, the isolated MP protein or portion thereof, is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: in the Sequence Listing, such as SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6) such that the protein or portion thereof maintains the ability to catalyze an enzymatic reaction involved in one or more pathways for the metabolism of an amino acid, a vitamin, a cofactor, a nutraceutical, a nucleotide, a nucleoside, or trehalose.

The invention also provides an isolated preparation of an MP protein. In preferred embodiments, the MP protein comprises an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing such as SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6). In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO of the Sequence Listing such as SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6) (encoded by an open reading frame set forth in a corresponding odd-numbered SEQ ID NO: of the Sequence Listing such as SEQ ID

NO:1, SEQ ID NO:3, or SEQ ID NO:5). In yet another embodiment, the protein is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing such as SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6). In other embodiments, the isolated MP protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing such as SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6) and is able to catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway either alone or in combination one or more MP proteins of the invention or any protein of the same or different metabolic pathway, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated MP protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, or is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to a nucleotide sequence of one of the even-numbered SEQ ID NOs set forth in the Sequence Listing. It is also preferred that the preferred forms of MP proteins also have one or more of the MP bioactivities described herein.

The MP polypeptide, or a biologically active portion thereof, can be operatively linked to a non-MP polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the MP protein alone. In other preferred embodiments, this fusion protein, when introduced into a *C. glutamicum* pathway for the metabolism of an amino acid, vitamin, cofactor, nutraceutical, results in increased yields and/or efficiency of production of a desired fine chemical from *C. glutamicum*. In particularly preferred embodiments, integration of this fusion protein

into an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway of a host cell modulates production of a desired compound from the cell.

In another aspect, the invention provides methods for screening molecules which modulate the activity of an MP protein, either by interacting with the protein itself or a substrate or binding partner of the MP protein, or by modulating the transcription or translation of an MP nucleic acid molecule of the invention.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing one or more vectors directing the expression of one or more MP nucleic acid molecules of the either alone or in combination one or more MP nucleic acid molecules of the invention or any nucleic acid molecule of the same or different metabolic pathway, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an MP nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3. In another preferred embodiment, the MP genes is the *metZ* gene (SEQ ID NO:1), *metC* gene (SEQ ID NO:3), or the gene designated as RXA00657 (SEQ ID NO:5) (see Table 1), alone or in combination with one or more MP nucleic acid molecules of the invention or with one or more genes involved in methionine and/or lysine metabolism. In yet another preferred embodiment, the fine chemical is an amino acid, *e.g.*, L-lysine and L-methionine.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates MP protein activity or MP nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, such that the yields or rate of production of a desired fine chemical by this microorganism is improved. The agent which modulates MP protein activity can be an agent which stimulates MP protein activity or MP nucleic acid expression. Examples of agents which stimulate MP protein activity or MP nucleic acid expression

include small molecules, active MP proteins, and nucleic acids encoding MP proteins that have been introduced into the cell. Examples of agents which inhibit MP activity or expression include small molecules and antisense MP nucleic acid molecules.

Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant MP gene into a cell, either alone or in combination one or more MP nucleic acid molecules of the invention or any nucleic acid molecule of the same or different metabolic pathway, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid are L-lysine and L-methionine. In another preferred embodiment, said gene is the *metZ* gene (SEQ ID NO:1), *metC* gene (SEQ ID NO:3), or the RXA00657 gene (SEQ ID NO:5), alone or in combination with one or more MP nucleic acid molecules of the invention or with one or more genes involved in methionine and/or lysine metabolism.

Detailed Description of the Invention

The present invention provides MP nucleic acid and protein molecules which are involved in the metabolism of certain fine chemicals in *Corynebacterium glutamicum*, including amino acids, *e.g.*, lysine and methionine, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (*e.g.*, where modulation of the activity of a lysine or methionine biosynthesis protein has a direct impact on the production or efficiency of production of lysine or methionine from that organism), or may have an indirect impact which nonetheless results in an increase of yield or efficiency of production of the desired compound (*e.g.*, where modulation of the activity of a nucleotide biosynthesis protein has an impact on the production of an organic acid or a fatty acid from the bacterium, perhaps due to improved growth or an increased supply of necessary co-factors, energy compounds, or precursor molecules). The MP molecules may be utilized

alone or in combination with other MP molecules of the invention, or in combination with other molecules involved in the same or a different metabolic pathway (*e.g.*, lysine or methione metabolism). In a preferred embodiment, the MP molecules are the *metZ* (SEQ ID NO:1), *metC* (SEQ ID NO:3), or RXA00657 (SEQ ID NO:5) nucleic acid molecules and the proteins encoded by these nucleic acid molecules (SEQ ID NO:2, SEQ ID NO.:4 and SEQ ID NO.:6, respectively). Aspects of the invention are further explicated below.

I. Fine Chemicals

The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described *e.g.* in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (*e.g.*, arachidonic acid), diols (*e.g.*, propane diol, and butane diol), carbohydrates (*e.g.*, hyaluronic acid and trehalose), aromatic compounds (*e.g.*, aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, polyketides (Cane *et al.* (1998) *Science* 282: 63-68), and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

A. *Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-

recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3rd edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosyntheses, are readily converted by simple biosynthetic pathways to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino aids – technical production and use, p. 466-502 in Rehm *et al.* (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E.(1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of α-ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain β-carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine.

The biosynthetic pathways leading to methionine have been studied in diverse organisms. The first step, acylation of homoserine, is common to all of the organisms, even though the source of the transferred acyl groups is different. *Escherichia coli* and the related species use succinyl-CoA (Michaeli, S. and Ron, E. Z. (1981) *Mol. Gen. Genet.* 182, 349-354), while *Saccharomyces cerevisiae* (Langin, T., *et al.* (1986) *Gene* 49, 283-293), *Brevibacterium flavum* (Miyajima, R. and Shiio, I. (1973) *J. Biochem.* 73, 1061-1068; Ozaki, H. and Shiio, I. (1982) *J. Biochem.* 91, 1163-1171), *C. glutamicum* (Park, S.-D., *et al.* (1998) *Mol. Cells* 8, 286-294), and *Leptospira meyeri* (Belfaiza, J. *et al.* (1998) 180, 250-255; Bourhy, P., *et al.* (1997) *J. Bacteriol.* 179, 4396-4398) use acetyl-CoA as the acyl donor. Formation of homocysteine from acylhomoserine can occur in two different ways. *E. coli* uses the transsulfuration pathway which is catalyzed by cystathionine γ-synthase (the product of *metB*) and cystathionine β-lyase

(the product of *metC*). *S. cerevisiae* (Cherest, H. and Surdin-Kerjan, Y. (1992) *Genetics* 130, 51-58), *B. flavum* (Ozaki, H. and Shiio, I. (1982) *J. Biochem.* 91, 1163-1171), *Pseudomonas aeruginosa* (Foglino, M., *et al.* (1995) *Microbiology* 141, 431-439), and *L. meyeri* (Belfaiza, J., *et al.* (1998) *J. Bacteriol.* 180, 250-255) utilize the direct sulfhydrylation pathway which is catalyzed by acylhomoserine sulfhydrylase. Unlike closely related *B. flavum* which uses only the direct sulfhydrylation pathway, enzyme activities of the transsulfuration pathway have been detected in the extracts of the *C. glutamicum* cells and the pathway has been proposed to be the route for methionine biosynthesis in the organism (Hwang, B-J., *et al.* (1999) *Mol. Cells* 9, 300-308; Kase, H. and Nakayama, K. (1974) *Agr. Biol. Chem.* 38, 2021-2030; Park, S.-D., *et al.* 1998) *Mol. Cells* 8, 286-294).

Although some genes involved in methionine biosynthesis in *C. glutamicum* have been isolated, information on the biosynthesis of methionine in *C. glutamicum* is still very limited. No genes other than *metA* and *metB* have been isolated from the organism. To understand the biosynthetic pathways leading to methionine in *C. glutamicum*, we have isolated and characterized the *metC* gene (SEQ ID NO:3) and the *metZ* (also called *metY*) gene (SEQ ID NO:1) of *C. glutamicum* (see Table 1).

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3rd ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3rd ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although

they are readily synthesized by other organisms, such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (*e.g.*, polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B_1) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B_2) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B_6' (*e.g.*, pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced

either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of β-alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to β-alanine and for the condensation to panthotenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B_5), pantetheine (and its derivatives) and coenzyme A.

Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the α-ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

Corrinoids (such as the cobalamines and particularly vitamin B_{12}) and porphyrines belong to a group of chemicals characterized by a tetrapyrole ring system. The biosynthesis of vitamin B_{12} is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B_6, pantothenate, and biotin. Only Vitamin B_{12} is produced solely by fermentation, due to the complexity of

its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (*i.e.*, AMP) or as coenzymes (*i.e.*, FAD and NAD).

Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (*e.g.* Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (*e.g.*, thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (*e.g.*, ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (*e.g.*, IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) Nucleotides and Related Compounds in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide

metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: Progress in Nucleic Acid Research and Molecular Biology, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

D. Trehalose Metabolism and Uses

Trehalose consists of two glucose molecules, bound in α, α-1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto *et al.*, (1998) U.S. Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech. Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) J. Japan 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

II. Elements and Methods of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as MP nucleic acid and protein molecules (see Table 1), which play a role in or function in one or more cellular metabolic pathways. In one embodiment, the MP molecules catalyze an enzymatic reaction involving one or more amino acid, *e.g.*, lysine or methionine, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways. In a preferred embodiment, the activity of one or more MP molecules of the present invention, alone or in combination with molecules involved in the same or different metabolic pathway (*e.g.*, methionine or lysine metabolism), in one or more *C. glutamicum* metabolic pathways for amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides or trehalose has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the MP molecules of the invention are modulated in activity, such that the *C. glutamicum* metabolic pathways in which the MP proteins of the invention are involved are modulated in efficiency or output, which either directly or indirectly modulates the production or efficiency of production of a desired fine chemical by *C. glutamicum*. In a preferred embodiment, the fine chemical is an amino acid, *e.g.*, lysine or methionine. In another preferred embodiment, the MP molecules are metZ, metY, and/or RXA00657 (see Table 1).

The language, "MP protein" or "MP polypeptide" includes proteins which play a role in, *e.g.*, catalyze an enzymatic reaction, in one or more amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside or trehalose metabolic pathways. Examples of MP proteins include those encoded by the MP genes set forth in Table 1 and by the odd-numbered SEQ ID NOs. The terms "MP gene" or "MP nucleic acid sequence" include nucleic acid sequences encoding an MP protein, which consist of a coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of MP genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (*e.g.*, kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes

the efficiency of the conversion of the carbon source into the product (*i.e.*, fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (*e.g.*, the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound.

The MP molecules of the present invention may be combined with one or more MP molecules of the invention or one or more molecules of the same or different metabolic pathway to increase the yield of a desired fine chemical. In a preferred embodiment, the fine chemical is an amino acid, *e.g.*, lysine or methionine. Alternatively, or in addition, a byproduct which is not desired may be reduced by combination or disruption of MP molecules or other metabolic molecules (*e.g.*, molecules involved in lysine or methionine metabolism). MP molecules combined with other molecules of the same or a different metabolic pathway may be altered in their nucleotide sequence and in the corresponding amino acid sequence to alter their activity under physiological conditions, which leads to an increase in productivity and/or yield of a desired fine chemical. In a further embodiment, an MP molecule in its original or in its above-described altered form may be combined with other molecules of the same or a different metabolic pathway which are altered in their nucleotide sequence in such a way that their activity is altered under physiological conditions which leads to an increase in productivity and/or yield of a desired fine chemical, *e.g.*, an amino acid such as methionine or lysine.

In another embodiment, the MP molecules of the invention, alone or in combination with one or more molecules of the same or different metabolic pathway, are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. Using recombinant genetic techniques, one or more of the biosynthetic or degradative enzymes of the invention for amino acids, *e.g.*, lysine or methionine, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose may be manipulated such that its function is modulated. For example, a biosynthetic enzyme may be improved in efficiency, or its allosteric control region destroyed such that feedback inhibition of production of the compound is prevented. Similarly, a degradative enzyme may be deleted or modified by substitution, deletion, or addition such that its degradative activity is lessened for the desired compound without impairing the viability of the cell. In each case, the overall yield or rate of production of one of these desired fine chemicals may be increased.

It is also possible that such alterations in the protein and nucleotide molecules of the invention may improve the production of other fine chemicals besides the amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose. Metabolism of any one compound is necessarily intertwined with other biosynthetic and degradative pathways within the cell, and necessary cofactors, intermediates, or substrates in one pathway are likely supplied or limited by another such pathway. Therefore, by modulating the activity of one or more of the proteins of the invention, the production or efficiency of activity of another fine chemical biosynthetic or degradative pathway may be impacted. For example, amino acids serve as the structural units of all proteins, yet may be present intracellularly in levels which are limiting for protein synthesis; therefore, by increasing the efficiency of production or the yields of one or more amino acids within the cell, proteins, such as biosynthetic or degradative proteins, may be more readily synthesized. Likewise, an alteration in a metabolic pathway enzyme such that a particular side reaction becomes more or less favored may result in the over- or under-production of one or more compounds which are utilized as intermediates or substrates for the production of a desired fine chemical.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* MP DNAs and the predicted amino acid sequences of the *C.*

glutamicum MP proteins are shown in the Sequence Listing as odd-numbered SEQ ID NOs and even-numbered SEQ ID NOs, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode metabolic pathway proteins, *e.g.*, proteins involved in the methionine or lysine metabolic pathways.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of the invention (*e.g.*, the sequence of an even-numbered SEQ ID NO of the Sequence Listing). As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, *e.g.*, the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to the selected amino acid sequence.

An MP protein of the invention, or a biologically active portion or fragment thereof, alone or in combination with one or more proteins of the same or different metabolic pathway, can catalyze an enzymatic reaction in one or more amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, or have one or more of the activities set forth in Table 1 (*e.g.*, metabolism of methionine or lysine biosynthesis).

Various aspects of the invention are described in further detail in the following subsections:

A. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode MP polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of MP-encoding nucleic acid (*e.g.*, MP DNA). As used herein, the term

"nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3'end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated MP nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g, a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a DNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having a nucleotide sequence of an odd-numbered SEQ ID NO of the Sequence Listing, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* MP DNA can be isolated from a *C. glutamicum* library using all or portion of one of the odd-numbered SEQ ID NO sequences of the Sequence Listing as a hybridization probe and standard hybridization techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the nucleic acid sequences of the invention (*e.g.*, an odd-numbered SEQ ID NO:) can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (*e.g.*, a nucleic acid molecule encompassing all or a portion of one of the nucleic acid sequences of the invention (*e.g.*, an odd-numbered SEQ ID NO of the

Sequence Listing) can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence). For example, mRNA can be isolated from normal endothelial cells (*e.g.*, by the guanidinium-thiocyanate extraction procedure of Chirgwin *et al.* (1979) *Biochemistry* 18: 5294-5299) and DNA can be prepared using reverse transcriptase (*e.g.*, Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in the Sequence Listing. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an MP nucleotide sequence can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in the Sequence Listing. The nucleic acid sequences of the invention, as set forth in the Sequence Listing, correspond to the *Corynebacterium glutamicum* MP DNAs of the invention. This DNA comprises sequences encoding MP proteins (*i.e.*, the "coding region", indicated in each odd-numbered SEQ ID NO: sequence in the Sequence Listing), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in each odd-numbered SEQ ID NO: in the Sequence Listing. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the nucleic acid sequences of the Sequence Listing.

For the purposes of this application, it will be understood that some of the MP nucleic acid and amino acid sequences set forth in the Sequence Listing have an identifying RXA, RXN, RXS, or RXC number having the designation "RXA", "RXN", "RXS", or "RXC" followed by 5 digits (*i.e.*, RXA, RXN, RXS, or RXC). Each of the nucleic acid sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA, RXN, RXS, or RXC designation to eliminate confusion. The recitation "one of the odd-numbered sequences of the Sequence Listing", then, refers to any of the nucleic acid sequences in the Sequence Listing, which may also be distinguished by their

differing RXA, RXN, RXS, or RXC designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is also set forth in the Sequence Listing, as an even-numbered SEQ ID NO: immediately following the corresponding nucleic acid sequence . For example, the coding region for RXA00115 is set forth in SEQ ID NO:69, while the amino acid sequence which it encodes is set forth as SEQ ID NO:70. The sequences of the nucleic acid molecules of the invention are identified by the same RXA, RXN, RXS, or RXC designations as the amino acid molecules which they encode, such that they can be readily correlated. For example, the amino acid sequences designated RXA00115, RXN00403, and RXS03158 are translations of the coding regions of the nucleotide sequences of nucleic acid molecules RXA00115, RXN00403, and RXS03158, respectively. The correspondence between the RXA, RXN, RXS, and RXC nucleotide and amino acid sequences of the invention and their assigned SEQ ID NOs is set forth in Table 1.

Several of the genes of the invention are "F-designated genes". An F-designated gene includes those genes set forth in Table 1 which have an 'F' in front of the RXA, RXN, RXS, or RXC designation. For example, SEQ ID NO:77, designated, as indicated on Table 1, as "F RXA00254", is an F-designated gene.

Also listed on Table 1 are the *metZ* (or *metY*) and *metC* genes (designated as SEQ ID NO:1 and SEQ ID NO:3, respectively. The corresponding amino acid sequence encoded by the *metZ* and *metC* genes are designated as SEQ ID NO:2 and SEQ ID NO:5, respectively.

In one embodiment, the nucleic acid molecules of the present invention are not intended to include those compiled in Table 2.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing), or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences of the invention is one which is sufficiently complementary to one of the nucleotide sequences shown in the Sequence Listing (*e.g.*, the sequence of an odd-numbered SEQ ID NO:) such that it can hybridize to one of the nucleotide sequences of the invention, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50%, 51%, 52%, 53%,

54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to a nucleotide sequence of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing), or a portion thereof. Ranges and identity values intermediate to the above-recited ranges, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to one of the nucleotide sequences of the invention, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of the sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an MP protein. The nucleotide sequences determined from the cloning of the MP genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning MP homologues in other cell types and organisms, as well as MP homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the nucleotide sequences of the invention (*e.g.*, a sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing), an anti-sense sequence of one of these sequences, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of the invention can be used in PCR reactions to clone MP homologues. Probes based on the MP nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-

factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an MP protein, such as by measuring a level of an MP-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting MP mRNA levels or determining whether a genomic MP gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO of the Sequence Listing) such that the protein or portion thereof maintains the ability to catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain as an amino acid residue in a sequence of one of the even-numbered SEQ ID NOs of the Sequence Listing) amino acid residues to an amino acid sequence of the invention such that the protein or portion thereof is able to catalyze an enzymatic reaction in a *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside or trehalose metabolic pathway. Protein members of such metabolic pathways, as described herein, function to catalyze the biosynthesis or degradation of one or more of: amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose. Examples of such activities are also described herein. Thus, "the function of an MP protein" contributes to the overall functioning of one or more such metabolic pathway and contributes, either directly or indirectly, to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of MP protein activities are set forth in Table 1.

In another embodiment, the protein is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing).

Portions of proteins encoded by the MP nucleic acid molecules of the invention are preferably biologically active portions of one of the MP proteins. As used herein, the term "biologically active portion of an MP protein" is intended to include a portion, *e.g.*, a domain/motif, of an MP protein that catalyzes an enzymatic reaction in one or more *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, or has an activity as set forth in Table 1. To determine whether an MP protein or a biologically active portion thereof can catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, an assay of enzymatic activity may be performed. Such assay methods are well known to those of ordinary skill in the art, as detailed in Example 8 of the Exemplification.

Additional nucleic acid fragments encoding biologically active portions of an MP protein can be prepared by isolating a portion of one of the amino acid sequences of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing), expressing the encoded portion of the MP protein or peptide (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the MP protein or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing) (and portions thereof) due to degeneracy of the genetic code and thus encode the same MP protein as that encoded by the nucleotide sequences of the invention. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in the Sequence Listing (*e.g.*, an even-numbered SEQ ID NO:). In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of the invention (encoded by an open reading frame shown in an odd-numbered SEQ ID NO: of the Sequence Listing).

It will be understood by one of ordinary skill in the art that in one embodiment the sequences of the invention are not meant to include the sequences of the prior art, such as those Genbank sequences set forth in Table 2, which was available prior to the present invention. In one embodiment, the invention includes nucleotide and amino acid sequences having a percent identity to a nucleotide or amino acid sequence of the

invention which is greater than that of a sequence of the prior art (*e.g.*, a Genbank sequence (or the protein encoded by such a sequence) set forth in Table 2). For example, the invention includes a nucleotide sequence which is greater than and/or at least 45% identical to the nucleotide sequence designated RXA00657 SEQ ID NO:5 One of ordinary skill in the art would be able to calculate the lower threshold of percent identity for any given sequence of the invention by examining the GAP-calculated percent identity scores set forth in Table 4 for each of the three top hits for the given sequence, and by subtracting the highest GAP-calculated percent identity from 100 percent. One of ordinary skill in the art will also appreciate that nucleic acid and amino acid sequences having percent identities greater than the lower threshold so calculated (*e.g.*, at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more identical) are also encompassed by the invention.

In addition to the *C. glutamicum* MP nucleotide sequences set forth in the Sequence Listing as odd-numbered SEQ ID NOs, it will be appreciated by one of ordinary skill in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of MP proteins may exist within a population (*e.g.*, the *C. glutamicum* population). Such genetic polymorphism in the MP gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an MP protein, preferably a *C. glutamicum* MP protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the MP gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in MP that are the result of natural variation and that do not alter the functional activity of MP proteins are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* MP DNA of the invention can be isolated based on their homology to the *C. glutamicum* MP nucleic acid disclosed herein using the *C. glutamicum* DNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in

another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of an odd-numbered SEQ ID NO: of the Sequence Listing. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to one of ordinary skill in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a nucleotide sequence of the invention corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* MP protein.

In addition to naturally-occurring variants of the MP sequence that may exist in the population, one of ordinary skill in the art will further appreciate that changes can be introduced by mutation into a nucleotide sequence of the invention, thereby leading to changes in the amino acid sequence of the encoded MP protein, without altering the functional ability of the MP protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a nucleotide sequence of the invention. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the MP proteins (*e.g.*, an even-numbered SEQ ID NO: of the Sequence Listing) without altering the activity of said MP protein, whereas an "essential" amino acid residue is required for MP protein activity. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-

conserved in the domain having MP activity) may not be essential for activity and thus are likely to be amenable to alteration without altering MP activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding MP proteins that contain changes in amino acid residues that are not essential for MP activity. Such MP proteins differ in amino acid sequence from a sequence of an even-numbered SEQ ID NO: of the Sequence Listing yet retain at least one of the MP activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of the invention and is capable of catalyzing an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, or has one or more activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% homologous to one of the amino acid sequences of the invention.

To determine the percent homology of two amino acid sequences (*e.g.*, one of the amino acid sequences of the invention and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (*e.g.*, one of the amino acid sequences of the invention) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (*e.g.*, a mutant form of the amino acid sequence), then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an MP protein homologous to a protein sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of

the Sequence Listing) can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of the invention such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the nucleotide sequences of the invention by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an MP protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an MP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an MP activity described herein to identify mutants that retain MP activity. Following mutagenesis of the nucleotide sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

In addition to the nucleic acid molecules encoding MP proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded DNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire MP coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is

antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MP protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the entire coding region of SEQ ID NO.:1 (*metZ*) comprises nucleotides 363 to 1673). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding MP. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding MP disclosed herein (*e.g.*, the sequences set forth as odd-numbered SEQ ID NOs in the Sequence Listing), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of MP mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of MP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of MP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine,

uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an MP protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes

(described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave MP mRNA transcripts to thereby inhibit translation of MP mRNA. A ribozyme having specificity for an MP-encoding nucleic acid can be designed based upon the nucleotide sequence of an MP DNA disclosed herein (*i.e.*, SEQ ID NO:1 (*metZ*). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an MP-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071 and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, MP mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, MP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an MP nucleotide sequence (*e.g.*, an MP promoter and/or enhancers) to form triple helical structures that prevent transcription of an MP gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

Another aspect of the invention pertains to combinations of genes involved in methionine and/or lysine metabolism and the use of to combinations of genes involved in methionine and/or lysine metabolism in the methods of the invention. Preferred combinations are the combination of *metZ* with *metC*, *metB* (encoding Cystathionine-Synthase), *metA* (encoding homoserine-O-acetyltransferase), *metE* (encoding Methionine Synthase), *metH* (encoding Methionine Synthase), *hom* (encoding homoserine dehydrogenase), *asd* (encoding aspartatesemialdehyd dehydrogenase), *lysC /ask* (encoding aspartokinase) and rxa00657 (herein designated as SEQ ID NO.:5), *dapA*, (gene encoding DIHYDRODIPICOLINATE SYNTHASE), *dapB* (gene encoding DIHYDRODIPICOLINATE REDUCTASE), *dapC* (gene encoding 2,3,4,5-tetrahydropyridine-2-carboxylate N—succinyltransferase), *dapD/argD* (gene encoding acetylornithine transaminase), *dapE* (gene encoding succinyldiaminopimelate desuccinylase), *dapF* (gene encoding diaminopimelate epimerase), *lysA* (gene encoding diaminopimelate decarboxylase), *ddh* (gene encoding diaminopimelate dehydrogenase), *lysE* (gene encoding for the lysine exporter), *lysG* (gene encoding for the exporter regulator), *hsk* (gene encoding homoserine kinase) as well as genes involved in anaplerotic reaction such as *ppc* (gene encoding phosphoenolpyruvate carboxylase),

ppcK (gene encoding phosphoenolpyruvate carboxykinase), *pycA* (gene encoding pyruvate carboxylase), *accD*, *accA*, *accB*, *accC* (genes encoding for subunits of acetyl-CoA-carboxylase), as well as genes of the pentose-phosphate pathway, gpdh genes encoding glucose-6-phophate-dehydrogenase, *opcA*, *pgdh* (gene encoding 6-phosphogluconate-dehydrogenase), *ta* (gene encoding transaldolase), *tk* (gene encoding gene encoding transketolase), *pgl* (gene encoding 6-PHOSPHOGLUCONO-LACTONASE), *rlpe* (gene encoding RIBULOSE-PHOSPHATE 3-EPIMERASE) *rpe* (gene encoding RIBOSE 5-PHOSPHATE EPIMERASE) or combinations of the above-mentioned genes of the pentose-phosphate-pathways, or other MP genes of the invention.

The genes may be altered in their nucleotide sequence and in the corresponding amino acid sequence resulting in derivatives in such a way that their activity is altered under physiological conditions which leads to an increase in productivity and/or yield of a desired fine chemical, *e.g.*, an amino acid such as methionine or lysine. One class of such alterations or derivatives is well known for the nucleotide sequence of the *ask* gene encoding aspartokinase. These alterations lead to removal of feed back inhibition by the amino acids lysine and threonine and subsequently to lysine overproduction. In a preferred embodiment the *metZ* gene or altered forms of the *metZ* gene are used in a *Corynebacterium* strain in combination with *ask*, *hom*, *metA* and *metH* or derivatives of these genes. In another preferred embodiment *metZ* or altered forms of the *metZ* gene are used in a *Corynebacterium* strain in combination with *ask*, *hom*, *metA* and *metE* or derivatives of these genes. In a more preferred embodiment, the gene combinations *metZ* or altered forms of the *metZ* gene are combined with *ask*, *hom*, *metA* and *metH* or derivatives of these genes, or *metZ* is combined with *ask*, *hom*, *metA* and *metE* or derivatives of these genes in a *Corynebacterium* strain and sulfur sources such as sulfates, thiosulfates, sulfites and also more reduced sulfur sources such as H_2S and sulfides and derivatives are used in the growth medium. Also, sulfur sources such as methyl mercaptan, methanesulfonic acid, thioglycolates, thiocyanates, thiourea, sulfur containing amino acids such as cysteine and other sulfur containing compounds can be used. Another aspect of the invention pertains to the use of the above mentioned gene combinations in a *Corynebacterium* strain which is, before or after introduction of the genes, mutagenized by radiation or by mutagenic chemicals well-known to one of ordinary skill in the art and selected for resistance against high concentrations of the fine

chemical of interest, *e.g.* lysine or methionine or analogues of the desired fine chemical such as the methionine analogues ethionine, methyl methionine, or others. In another embodiment, the gene combinations mentioned above can be expressed in a *Corynebacterium* strain having particular gene disruptions. Preferred are gene disruptions that encode proteins that favor carbon flux to undesired metabolites. Where methionine is the desired fine chemical the formation of lysine may be unfavorable. In such a case the combination of the above mentioned genes should proceed in a *Corynebacterium* strain bearing a gene disruption of the *lysA* gene (encoding diaminopimelate decarboxylase) or the *ddh* gene (encoding the meso-diaminopimelate dehydrogenase catalysing the conversion of tetrahydropicolinate to meso-diaominopimelate). In a preferred embodiment, a favorable combination of the above-mentioned genes are all altered in such a way that their gene products are not feed back inhibited by end products or metabolites of the biosynthetic pathway leading to the desired fine chemical. In the case that the desired fine chemical is methionine, the gene combinations may be expressed in a strain previously treated with mutagenic agents or radiation and selected for the above-mentioned resistance. Additionally, the strain should be grown in a growth medium containing one or more of the above mentioned sulfur sources.

In another embodiment of the invention, a gene was identified from the genome of *Corynebacterium glutamicum* as a gene coding for a hypothetical transcriptional regulatory protein. This gene is described as RXA00657. The nucleotide sequence of RXA00657 corresponds to SEQ ID NO:5. The amino acid sequence of RXA00657 corresponds to SEQ ID NO:6. It was found that when the RXA00657 gene, as well as upstream and downstream regulatory regions described in the examples, was cloned into a vector capable of replicating in *Corynebacterium glutamicum* and transformed and expressed in a lysine producing strain such as ATCC13286, that this strain produced more lysine compared to the strain transformed with the same plasmid lacking the aforementioned nucleotide fragment RXA00657. In addition to the observation that the lysine titer was increased in the mentioned strain, the selectivity determined by the molar amount of lysine produced compared to the molar amount of sucrose consumed was increased (see Example 14). Overexpression of RXA00657 in combination with the overexpression of other genes either directly involved in the lysine specific pathway

such as *lysC*, *dapA*, *dapB*, *dapC*, *dapD*, *dapF*, *ddh*, *lysE*, *lysG*, and *lysR* results in an increase in the production of lysine compared to RXA00657 alone.

B. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an MP protein (or a portion thereof) or combinations of genes wherein at least one gene encodes for an MP protein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory

sequence" is intended to include promoters, repressor binding sites, activator binding sites, enhancers and other expression control elements (*e.g.*, terminators, polyadenylation signals, or other elements of mRNA secondary structure). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells. Preferred regulatory sequences are, for example, promoters such as cos-, tac-, trp-, tet-, trp-tet-, lpp-, lac-, lpp-lac-, $lacI^q$-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, arny, SPO2, λ-P_R- or λ P_L, which are used preferably in bacteria. Additional regulatory sequences are, for example, promoters from yeasts and fungi, such as ADC1, MFα, AC, P-60, CYC1, GAPDH, TEF, rp28, ADH, promoters from plants such as CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos or ubiquitin- or phaseolin-promoters. It is also possible to use artificial promoters. It will be appreciated by one of ordinary skill in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, MP proteins, mutant forms of MP proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of MP proteins in prokaryotic or eukaryotic cells. For example, MP genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. *et al.* (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. *et al.* (1991) "Heterologous gene expression in filamentous fungi" in: More Gene Manipulations in Fungi, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy, J.F. *et al.*, eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens* –mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian cells.

Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the MP protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant MP protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHS1, pHS2, pPLc236, pMBL24, pLG200, pUR290, pIN-III113-B1, λgt11, pBdCl, and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89; and Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018). Target gene expression from the pTrc vector relies on host RNA polymerase

transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter. For transformation of other varieties of bacteria, appropriate vectors may be selected. For example, the plasmids pIJ101, pIJ364, pIJ702 and pIJ361 are known to be useful in transforming Streptomyces, while plasmids pUB110, pC194, or pBD214 are suited for transformation of Bacillus species. Several plasmids of use in the transfer of genetic information into Corynebacterium include pHM1519, pBL1, pSA77, or pAJ667 (Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018).

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada *et al.* (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the MP protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), , 2 μ, pAG-1, Yep6, Yep13, pEMBLYe23, pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, *et al.*, eds., p. 1-28, Cambridge University Press: Cambridge, and Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York (IBSN 0 444 904018).

Alternatively, the MP proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of

proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In another embodiment, the MP proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (*e.g.*, the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan, M.W. (1984) "Binary *Agrobacterium* vectors for plant transformation", *Nucl. Acid. Res.* 12: 8711-8721, and include pLGV23, pGHlac+, pBIN19, pAK2004, and pDH51 (Pouwels *et al.*, eds. (1985) Cloning Vectors. Elsevier: New York IBSN 0 444 904018).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477),

pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to MP mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an MP protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those of ordinary skill in the art. Microorganisms

related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, linear DNA or RNA (*e.g.*, a linearized vector or a gene construct alone without a vector) or nucleic acid in the form of a vector (*e.g.*, a plasmid, phage, phasmid, phagemid, transposon or other DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an MP protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an MP gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the MP gene. Preferably, this MP gene is a *Corynebacterium glutamicum* MP gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous MP gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively,

the vector can be designed such that, upon homologous recombination, the endogenous MP gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous MP protein). In the homologous recombination vector, the altered portion of the MP gene is flanked at its 5' and 3' ends by additional nucleic acid of the MP gene to allow for homologous recombination to occur between the exogenous MP gene carried by the vector and an endogenous MP gene in a microorganism. The additional flanking MP nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see *e.g.*, Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (*e.g.*, by electroporation) and cells in which the introduced MP gene has homologously recombined with the endogenous MP gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an MP gene on a vector placing it under control of the lac operon permits expression of the MP gene only in the presence of IPTG. Such regulatory systems are well known in the art.

In another embodiment, an endogenous MP gene in a host cell is disrupted (*e.g.*, by homologous recombination or other genetic means known in the art) such that expression of its protein product does not occur. In another embodiment, an endogenous or introduced MP gene in a host cell has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional MP protein. In still another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or inducer) of an MP gene in a microorganism has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of the MP gene is modulated. One of ordinary skill in the art will appreciate that host cells containing more than one of the described MP gene and protein modifications may be readily produced using the methods of the invention, and are meant to be included in the present invention.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an MP protein. Accordingly, the invention

further provides methods for producing MP proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an MP protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered MP protein) in a suitable medium until MP protein is produced. In another embodiment, the method further comprises isolating MP proteins from the medium or the host cell.

C. Isolated MP Proteins

Another aspect of the invention pertains to isolated MP proteins, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of MP protein in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of MP protein having less than about 30% (by dry weight) of non-MP protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-MP protein, still more preferably less than about 10% of non-MP protein, and most preferably less than about 5% non-MP protein. When the MP protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of MP protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of MP protein having less than about 30% (by dry weight) of chemical precursors or non-MP chemicals, more preferably less than about 20% chemical precursors or non-MP chemicals, still more preferably less than about 10% chemical precursors or non-MP chemicals, and most preferably less than about 5% chemical precursors or non-MP chemicals. In preferred embodiments, isolated proteins or

biologically active portions thereof lack contaminating proteins from the same organism from which the MP protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* MP protein in a microorganism such as *C. glutamicum*.

An isolated MP protein or a portion thereof of the invention can catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) such that the protein or portion thereof maintains the ability to catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an MP protein of the invention has an amino acid sequence set forth as an even-numbered SEQ ID NO: of the Sequence Listing. In yet another preferred embodiment, the MP protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to a nucleotide sequence of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing). In still another preferred embodiment, the MP protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to one of the nucleic acid sequences of the invention, or a portion thereof. Ranges and identity values intermediate to the above-recited values, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. The preferred MP proteins of the present invention also preferably possess at least one of the MP activities described herein. For example, a preferred MP protein of the present invention includes an amino acid sequence encoded by a

nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to a nucleotide sequence of the invention, and which can catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, or which has one or more of the activities set forth in Table 1.

In other embodiments, the MP protein is substantially homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) and retains the functional activity of the protein of one of the amino acid sequences of the invention yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the MP protein is a protein which comprises an amino acid sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99%, 99.7% or more homologous to an entire amino acid sequence of the invention and which has at least one of the MP activities described herein. Ranges and identity values intermediate to the above-recited values, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of the invention.

Biologically active portions of an MP protein include peptides comprising amino acid sequences derived from the amino acid sequence of an MP protein, *e.g.*, an amino acid sequence of an even-numbered SEQ ID NO: of the Sequence Listing or the amino acid sequence of a protein homologous to an MP protein, which include fewer amino acids than a full length MP protein or the full length protein which is homologous to an MP protein, and exhibit at least one activity of an MP protein. Typically, biologically active portions (peptides, *e.g.*, peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an MP protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant

techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an MP protein include one or more selected domains/motifs or portions thereof having biological activity.

MP proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the MP protein is expressed in the host cell. The MP protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an MP protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native MP protein can be isolated from cells (*e.g.*, endothelial cells), for example using an anti-MP antibody, which can be produced by standard techniques utilizing an MP protein or fragment thereof of this invention.

The invention also provides MP chimeric or fusion proteins. As used herein, an MP "chimeric protein" or "fusion protein" comprises an MP polypeptide operatively linked to a non-MP polypeptide. An "MP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to MP, whereas a "non-MP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the MP protein, *e.g.*, a protein which is different from the MP protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the MP polypeptide and the non-MP polypeptide are fused in-frame to each other. The non-MP polypeptide can be fused to the N-terminus or C-terminus of the MP polypeptide. For example, in one embodiment the fusion protein is a GST-MP fusion protein in which the MP sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant MP proteins. In another embodiment, the fusion protein is an MP protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of an MP protein can be increased through use of a heterologous signal sequence.

Preferably, an MP chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended

termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An MP-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the MP protein.

Homologues of the MP protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the MP protein. As used herein, the term "homologue" refers to a variant form of the MP protein which acts as an agonist or antagonist of the activity of the MP protein. An agonist of the MP protein can retain substantially the same, or a subset, of the biological activities of the MP protein. An antagonist of the MP protein can inhibit one or more of the activities of the naturally occurring form of the MP protein, by, for example, competitively binding to a downstream or upstream member of the MP cascade which includes the MP protein. Thus, the *C. glutamicum* MP protein and homologues thereof of the present invention may modulate the activity of one or more metabolic pathways in which MP proteins play a role in this microorganism.

In an alternative embodiment, homologues of the MP protein can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the MP protein for MP protein agonist or antagonist activity. In one embodiment, a variegated library of MP variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of MP variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential MP sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of MP sequences therein. There are a variety of methods which can be used to produce libraries of potential MP

homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential MP sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the MP protein coding can be used to generate a variegated population of MP fragments for screening and subsequent selection of homologues of an MP protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an MP coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the MP protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of MP homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify MP homologues (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In another embodiment, cell based assays can be exploited to analyze a variegated MP library, using methods well known in the art.

D. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of MP protein regions required for function; modulation of an MP protein activity; modulation of the activity of an MP pathway; and modulation of cellular production of a desired compound, such as a fine chemical.

The MP nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is not pathogenic to humans, it is related to species which are human pathogens, such as *Corynebacterium diphtheriae*. *Corynebacterium diphtheriae* is the causative agent of diphtheria, a rapidly developing, acute, febrile infection which involves both local and systemic pathology. In this disease, a local lesion develops in the upper respiratory tract and involves necrotic injury to epithelial cells; the bacilli secrete toxin which is disseminated through this lesion to distal susceptible tissues of the body. Degenerative changes brought about by the inhibition of protein synthesis in these tissues, which include heart, muscle, peripheral nerves, adrenals, kidneys, liver and spleen, result in the systemic pathology of the disease. Diphtheria continues to have high incidence in many parts of the world, including Africa, Asia, Eastern Europe and the independent states of the former Soviet Union. An ongoing epidemic of diphtheria in the latter two regions has resulted in at least 5,000 deaths since 1990.

In one embodiment, the invention provides a method of identifying the presence or activity of *Cornyebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (*e.g.*, the sequences set forth as odd-numbered or even-numbered SEQ ID NOs, respectively, in the Sequence Listing) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject. *C. glutamicum* and *C. diphtheriae* are related bacteria, and many of the nucleic acid and protein molecules in *C. glutamicum* are homologous to *C. diphtheriae* nucleic acid and protein molecules, and can therefore be used to detect *C. diphtheriae* in a subject.

The nucleic acid and protein molecules of the invention may also serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

The MP nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may

give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

Manipulation of the MP nucleic acid molecules of the invention may result in the production of MP proteins having functional differences from the wild-type MP proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

The invention also provides methods for screening molecules which modulate the activity of an MP protein, either by interacting with the protein itself or a substrate or binding partner of the MP protein, or by modulating the transcription or translation of an MP nucleic acid molecule of the invention. In such methods, a microorganism expressing one or more MP proteins of the invention is contacted with one or more test compounds, and the effect of each test compound on the activity or level of expression of the MP protein is assessed.

When the desired fine chemical to be isolated from large-scale fermentative culture of *C. glutamicum* is an amino acid, a vitamin, a cofactor, a nutraceutical, a nucleotide, a nucleoside, or trehalose, modulation of the activity or efficiency of activity of one or more of the proteins of the invention by recombinant genetic mechanisms may directly impact the production of one of these fine chemicals. For example, in the case of an enzyme in a biosynthetic pathway for a desired amino acid, improvement in efficiency or activity of the enzyme (including the presence of multiple copies of the gene) should lead to an increased production or efficiency of production of that desired amino acid. In the case of an enzyme in a biosynthetic pathway for an amino acid whose synthesis is in competition with the synthesis of a desired amino acid, any decrease in the efficiency or activity of this enzyme (including deletion of the gene) should result in an increase in production or efficiency of production of the desired amino acid, due to decreased competition for intermediate compounds and/or energy. In the case of an enzyme in a degradation pathway for a desired amino acid, any decrease in efficiency or activity of the enzyme should result in a greater yield or efficiency of production of the desired product due to a decrease in its degradation. Lastly, mutagenesis of an enzyme involved in the biosynthesis of a desired amino acid such that this enzyme is no longer is capable of feedback inhibition should result in increased yields or efficiency of production of the desired amino acid. The same should apply to the biosynthetic and

degradative enzymes of the invention involved in the metabolism of vitamins, cofactors, nutraceuticals, nucleotides, nucleosides and trehalose.

Similarly, when the desired fine chemical is not one of the aforementioned compounds, the modulation of activity of one of the proteins of the invention may still impact the yield and/or efficiency of production of the compound from large-scale culture of *C. glutamicum*. The metabolic pathways of any organism are closely interconnected; the intermediate used by one pathway is often supplied by a different pathway. Enzyme expression and function may be regulated based on the cellular levels of a compound from a different metabolic process, and the cellular levels of molecules necessary for basic growth, such as amino acids and nucleotides, may critically affect the viability of the microorganism in large-scale culture. Thus, modulation of an amino acid biosynthesis enzyme, for example, such that it is no longer responsive to feedback inhibition or such that it is improved in efficiency or turnover may result in increased cellular levels of one or more amino acids. In turn, this increased pool of amino acids provides not only an increased supply of molecules necessary for protein synthesis, but also of molecules which are utilized as intermediates and precursors in a number of other biosynthetic pathways. If a particular amino acid had been limiting in the cell, its increased production might increase the ability of the cell to perform numerous other metabolic reactions, as well as enabling the cell to more efficiently produce proteins of all kinds, possibly increasing the overall growth rate or survival ability of the cell in large scale culture. Increased viability improves the number of cells capable of producing the desired fine chemical in fermentative culture, thereby increasing the yield of this compound. Similar processes are possible by the modulation of activity of a degradative enzyme of the invention such that the enzyme no longer catalyzes, or catalyzes less efficiently, the degradation of a cellular compound which is important for the biosynthesis of a desired compound, or which will enable the cell to grow and reproduce more efficiently in large-scale culture. It should be emphasized that optimizing the degradative activity or decreasing the biosynthetic activity of certain molecules of the invention may also have a beneficial effect on the production of certain fine chemicals from *C. glutamicum*. For example, by decreasing the efficiency of activity of a biosynthetic enzyme in a pathway which competes with the biosynthetic pathway of a desired compound for one or more intermediates, more of those intermediates should be available for conversion to the desired product. A similar

situation may call for the improvement of degradative ability or efficiency of one or more proteins of the invention.

This aforementioned list of mutagenesis strategies for MP proteins to result in increased yields of a desired compound is not meant to be limiting; variations on these mutagenesis strategies will be readily apparent to one of ordinary skill in the art. By these mechanisms, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated MP nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any natural product of *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention. Preferred compounds to be produced by *Corynebacterium glutamicum* strains are the amino acids L-lysine and L-methionine.

In one embodiment, the *metC* gene encoding cystathionine β-lyase, the third enzyme in the methionine biosynthetic pathway, was isolated from *Corynebacterium glutamicum*. The translational product of the gene showed no significant homology with that of *metC* gene from other organisms. Introduction of the plasmid containing the *metC* gene into *C. glutamicum* resulted in a 5-fold increase in the activity of cystathionine β-lyase. The protein product, now designated MetC (corresponding to SEQ ID NO:4), which encodes a protein product of 35,574 Daltons and consists of 325 amino acids, is identical to the previously reported *aecD* gene (Rossol, I. and Puhler, A. (1992) *J. Bacteriology* 174, 2968-2977) except the existence of two different amino acids. Like *aecD* gene, when present in multiple copies, *metC* gene conferred resistance to *S*-(β-aminoethyl)-cysteine which is a toxic lysine analog. However, genetic and biochemical evidences suggest that the natural activity of *metC* gene product is to mediate methionine biosynthesis in *C. glutamicum*. Mutant strains of *metC* were constructed and the strains showed methionine prototrophy. The mutant strains completely lost their ability to show resistance to *S*-(γ-aminoethyl)-cysteine. These results show that, in addition to the transsulfuration, which is another biosynthetic pathway, the direct sulfhydrylation pathway is functional in *C. glutamicum* as a parallel biosynthetic route for methionine.

In yet another embodiment, it is also shown that the additional sulfhydrylation pathway is catalyzed by *O*-acetylhomoserine sulfhydrylase. The presence of the pathway is demonstrated by the isolation of the corresponding *metZ* (or *metY*) gene and enzyme (corresponding to SEQ ID NO:1 and SEQ ID NO:2, respectively). Among the eukaryotes, fungi and yeast species have been reported to have both the transsulfuration and direct sulfhydrylation pathway. Thus far, no prokaryotic organism which possesses both pathways has been found. Unlike *E. coli* which only possesses single biosynthetic route for lysine, *C. glutamicum* possesses two parallel biosynthetic pathways for the amino acid. The biosynthetic pathway for methionine in *C. glutamicum* is analogous to that of lysine in that aspect.

The gene *metZ* is located in the upstream region of *metA*, which is the gene encoding the enzyme catalysing the first step of methionine biosynthesis (Park, S.-D., *et al.* (1998) *Mol. Cells* 8, 286-294). Regions upstream and downstream of *metA* were sequenced to identify other *met* genes. It appears that *metZ* and *metA* form an operon. Expression of the genes encoding MetA and MetZ leads to overproduction of the corresponding polypeptides.

Surprisingly, *metZ* clones can complement methionine auxotrophic *Escherichia coli metB* mutant strains. This shows that the protein product of *metZ* catalyzes a step that can bypass the step catalyzed by the protein product of *metB*. *MetZ* was also disrupted and the mutant strain showed methionine prototrophy. *Corynebacterium glutamicum metB* and *metZ* double mutants were also constructed. The double mutant is auxotrophic for methionine. Thus, *metZ* encodes a protein catalysing the reaction from O-Acetyl-Homoserine to Homocysteine, which is one step in the sulfhydrylation pathway of methionine biosynthesis. *Corynebacterium glutamicum* contains both the transsulfuration and the sulfhydrylation pathway of methionine biosynthesis.

Introduction of *metZ* into *C. glutamicum* resulted in the expression of a 47,000 Dalton protein. Combined introduction of *metZ* and *metA* in *C. glutamicum* resulted in the appearance of *metA* and *metZ* proteins as shown by gel electrophoresis. If the Corynebacterium strain is a lysine overproducer, introduction of a plasmid containing *metZ* and *metA* resulted in a lower lysine titer but accumulation of homocysteine and methionine is detected.

In another embodiment *metZ* and *metA* were introduced into *Corynebacterium glutamicum* strains together with the *hom* gene, encoding the homoserine dehydrogenase, catalysing the conversion from aspartate semialdehyde to homoserine. Different *hom* genes from different organisms were chosen for this experiment. The *Corynebacterium glutamicum hom* gene can be used as well as *hom* genes from other procaryotes like *Escherichia coli* or *Bacillus subtilis* or the *hom* gene of eukaryotes such as *Saccharomyces cerevisiae, Shizosaccharomyces pombe, Ashbya gossypii* or algae, higher plants or animals. It may be that the *hom* gene is insensitive against feed back inhibition mediated by any metabolites that occur in the biosynthetic routes of the amino acids of the aspartate family, like aspatrate, lysine, threonine or methionine. Such metabolites are for example aspartate, lysine, methionine, threonine, aspartyl-phosphate, aspartate semialdehyd, homoserine, cystathionine, homocysteine or any other metabolite that occurs in this biosynthetic routes. In addition to the metabolites, the homoserine dehydrogenase may be insensitive against inhibition by analogues of all those metabolites or even against other compounds involved in this metabolism as there are other amino acids like cysteine or cofactors like vitamin B12 and all of its derivatives and S-adenosylmethionine and its metabolites and derivatives and analogues. The insensitivity of the homoserine dehydrogenase against all these, a part of these or only one of these compounds may either be its natural attitude or it may be the result from one or more mutations that resulted from classical mutation and selection using chemicals or irradiation or other mutagens. The mutations could also be introduced into the *hom* gene using gene technology, for example the introduction of site specific point mutations or by any method aforementioned for the MP or MP encoding DNA-sequences.

When a *hom* gene was combined with the *metZ* and *metA* genes and introduced into a *Corynebacterium glutamicum* strain that is a lysine overproducer, lysine accumulation was reduced and homocysteine and methionine accumulation was enhanced. A further enhancement of homocysteine and methionine concentrations can be achieved, if a lysine overproducing *Corynebacterium glutamicum* strain is used and a disruption of the *ddh* gene or the *lysA* gene was introduced prior to the transformation with DNA containing a *hom* gene and *metZ* and *metA* in combination. The overproduction of homocysteine and methionine was possible using different sulfur sources. Sulfates, thiosulfates, sulfites and also more reduced sulfur sources like H_2S and sulfides and derivatives could be used. Also, organic sulfur sources like methyl mercaptan,

thioglycolates, thiocyanates, thiourea, sulfur containing amino acids like cysteine and other sulfur containing compounds can be used to achieve homocysteine and methionine overproduction.

In another embodiment, the *metC* gene was introduced into *a Corynebacterium glutamicum* strain using aforementioned methods. The *metC* gene can be transformed into the strain in combination with other genes like *metB*, *metA* and *metA*. The *hom* gene can also be added. When the *hom* gene, the *met C*, *metA* and *metB* genes were combined on a vector and introduced into a *Corynebacterium glutamicum* strain, homocysteine and methionine overproduction was achieved. The overproduction of homocysteine and methionine was possible using different sulfur sources. Sulfates, thiosulfates, sulfites and also more reduced sulfur sources like H_2S and sulfides and derivatives could be used. Also, organic sulfur sources like methyl mercaptan, thioglycolates, thiocyanates, thiourea, sulfur containing amino acids like cysteine and other sulfur containing compounds can be used to achieve homocysteine and methionine overproduction.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, published patent applications, Tables, and the sequence listing cited throughout this application are hereby incorporated by reference.

Exemplification

Example 1: Preparation of total genomic DNA of *Corynebacterium glutamicum* ATCC13032

A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30°C with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-I (5% of the original volume of the culture — all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-I: 140.34 g/l sucrose, 2.46 g/l $MgSO_4$ x $7H_2O$, 10 ml/l KH_2PO_4 solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l $(NH_4)_2SO_4$, 1 g/l NaCl, 2 g/l $MgSO_4$ x $7H_2O$, 0.2 g/l $CaCl_2$, 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l $FeSO_4$

x H_2O, 10 mg/l $ZnSO_4$ x 7 H_2O, 3 mg/l $MnCl_2$ x 4 H_2O, 30 mg/l H_3BO_3 20 mg/l $CoCl_2$ x 6 H_2O, 1 mg/l $NiCl_2$ x 6 H_2O, 3 mg/l Na_2MoO_4 x 2 H_2O, 500 mg/l complexing agent (EDTA or critic acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-panthothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37°C, the cell wall was degraded and the resulting protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-I and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 μg/ml, the suspension is incubated for ca.18 h at 37°C. The DNA was purified by extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20°C and a 30 min centrifugation at 12,000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20 μg/ml RNaseA and dialysed at 4°C against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20°C, the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

Example 2: Construction of genomic libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.

Using DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (*see e.g.*, Sambrook, J. *et al.* (1989) "Molecular Cloning : A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J.G. (1979) *Proc. Natl. Acad. Sci. USA*, 75:3737-3741); pACYC177 (Change & Cohen (1978) *J. Bacteriol* 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T.J., Rosenthal A. and Waterson, R.H. (1987) *Gene* 53:283-286. Gene libraries specifically for use in *C. glutamicum* may be constructed using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

For the isolation of *metC* clones, *E. coli* JE6839 cells were transformed with the library DNA and plated onto the M9 minimal medium containing ampicillin and appropriate supplements. The plates were incubated at 37°C for 5 days. Colonies were isolated and screened for the plasmid content. The complete nucleotide sequence of the isolated *metC* gene was determined by methods well-known to one of ordinary skill in the art.

Example 3: DNA Sequencing and Computational Functional Analysis

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see *e.g.*, Fleischman, R.D. *et al.* (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., *Science*, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' (SEQ ID NO:123) or 5'-GTAAAACGACGGCCAGT-3'(SEQ ID NO.:124).

Example 4: *In vivo* Mutagenesis

In vivo mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (*e.g. Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (*e.g.*, mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those of ordinary skill in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34.

Example 5: DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum*

Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as *e.g.*, pHM1519 or pBL1) which replicate autonomously (for review see, *e.g.*, Martin, J.F. *et al.* (1987) *Biotechnology*, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E.L. (1987) "From Genes to Clones — Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli and C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see *e.g.*, Yoshihama, M. *et al.* (1985) *J. Bacteriol.* 162:591-597, Martin J.F. *et al.* (1987) *Biotechnology*, 5:137-146 and Eikmanns, B.J. *et al.* (1991) *Gene*, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. *et al.* (1984) *J. Bacteriol.* 159306-311), electroporation (Liebl, E. *et al.* (1989) *FEMS Microbiol. Letters*, 53:399-303) and in cases where special vectors are used, also by conjugation (as described *e.g.* in Schäfer, A *et al.* (1990) *J. Bacteriol.* 172:1663-1666). It is also possible to transfer the shuttle vectors for *C. glutamicum* to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) *J. Mol. Biol.* 166:1-19).

Genes may be overexpressed in *C. glutamicum* strains using plasmids which comprise pCG1 (U.S. Patent No. 4,617,267) or fragments thereof, and optionally the

gene for kanamycin resistance from TN903 (Grindley, N.D. and Joyce, C.M. (1980) *Proc. Natl. Acad. Sci. USA* 77(12): 7176-7180). In addition, genes may be overexpressed in *C. glutamicum* strains using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

Aside from the use of replicative plasmids, gene overexpression can also be achieved by integration into the genome. Genomic integration in *C. glutamicum* or other Corynebacterium or Brevibacterium species may be accomplished by well-known methods, such as homologous recombination with genomic region(s), restriction endonuclease mediated integration (REMI) (see, *e.g.*, DE Patent 19823834), or through the use of transposons. It is also possible to modulate the activity of a gene of interest by modifying the regulatory regions (*e.g.*, a promoter, a repressor, and/or an enhancer) by sequence modification, insertion, or deletion using site-directed methods (such as homologous recombination) or methods based on random events (such as transposon mutagenesis or REMI). Nucleic acid sequences which function as transcriptional terminators may also be inserted 3' to the coding region of one or more genes of the invention; such terminators are well-known in the art and are described, for example, in Winnacker, E.L. (1987) From Genes to Clones – Introduction to Gene Technology. VCH: Weinheim.

Example 6: Assessment of the Expression of the Mutant Protein

Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel *et al.* (1988) Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by

several methods, all well-known in the art, such as that described in Bormann, E.R. *et al.* (1992) *Mol. Microbiol.* 6: 317-326.

To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as SDS-acrylamide gel electrophoresis, were employed. The overproduction of *metC* and *metZ* in combination with *metA* in *Corynebacterium glutamicum* was demonstrated by this method. Western blot may also be employed (see, for example, Ausubel *et al.* (1988) Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

Example 7: Growth of Escherichia coli and Genetically Modified Corynebacterium glutamicum — Media and Culture Conditions

E. coli strains are routinely grown in MB and LB broth, respectively (Follettie, M. T., *et al.* (1993) *J. Bacteriol.* 175, 4096-4103). Minimal media for *E. coli* is M9 and modified MCGC (Yoshihama, M., *et al.* (1985) *J. Bacteriol.* 162, 591-507). Glucose was added to a final concentration of 1%. Antibiotics were added in the following amounts (micrograms per milliliter): ampicillin, 50; kanamycin, 25; nalidixic acid, 25. Amino acids, vitamins, and other supplements were added in the following amounts: methionine, 9.3 mM; arginine, 9.3 mM; histidine, 9.3 mM; thiamine, 0.05 mM. *E. coli* cells were routinely grown at 37°C, respectively.

Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for Corynebacteria are both well-known and readily available (Lieb *et al.* (1989) *Appl. Microbiol. Biotechnol.*, 32:205-210; von der Osten *et al.* (1998) Biotechnology Letters, 11:11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: The Procaryotes, Volume II, Balows, A. *et al.*, eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace elements. Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose serve as

very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts, such as NH_4Cl or $(NH_4)_2SO_4$, NH_4OH, nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

The overproduction of sulfur containing amino acids like homocysteine and methionine was made possible using different sulfur sources. Sulfates, thiosulfates, sulfites and also more reduced sulfur sources like H_2S and sulfides and derivatives can be used. Also, organic sulfur sources like methyl mercaptan, thioglycolates, thiocyanates, thiourea, sulfur containing amino acids like cysteine and other sulfur containing compounds can be used to achieve homocysteine and methionine overproduction

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (*eds.* P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3). It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFCO) or others.

All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121°C) or by sterile filtration. The components can either be sterilized together or, if

necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature should be in a range between 15°C and 45°C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH_4OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the micro-organisms, the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100 – 300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD_{600} of O.5 – 1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2.5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30°C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

Example 8 – *In vitro* Analysis of the Function of Mutant Proteins

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one of ordinary skill in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3^{rd} ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2^{nd} ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3^{rd} ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

Cell extracts from *Corynebacterium glutamicum* were prepared as described previously (Park, S.-D., *et al.* (1998) *Mol. Cells* 8, 286-294). Cystathionine β-lyase was assayed as follows. The assay mixture contained 100 mM Tris-HCl (pH8.5), 0.1 mM NADH, 1 mM *L*-cystathionine, 5 units of *L*-lactate dehydrogenase, and appropriate amounts of crude extract. Optical changes were monitored at 340 nm. Assay for *S*-([illegible]-aminoethyl)-cysteine (AEC) resistance was carried out as described in Rossol, I. and Pühler, A. (1992) *J. Bacteriol.* 174, 2968-77. The results of cystathionin β-lyase assays from extracts of different *Corynebacterium glutamicum* strains as well as results of AEC resistance assays of the same strain are summarized in Table 5, below.

Table 5. Expression of cystathionine β-lyase [a]

Strains	Properties	Activity (nmol min^{-1} mg^{-1})	Growth on MM[b]	Resistance to AEC[c]
C. glutamicum ASO19E12	-	146	+	+
C. glutamicum ASO19E12/pMT1	Empty vector	145	+	+
C. glutamicum ASO19E12/pSL173	*metC* clone	797	+	++
C. glutamicum HL457	*metC* mutant[d]	19	+	-
C. glutamicum HL459	*metC* mutant[d]	23	+	-
E. coli JE6839	*metC* mutant	21	-	*ND*[e]

[a] The enzyme was induced by growth to the stationary phase on the minimal medium containing 1% glucose. Cells were harvested, disrupted, and assayed for the activity as described in the Materials and Methods.
[b] MCGC minimal media was used. Growth was monitored on plates.
[c] Cells were grown on plates containing 40 mM *S*-(β-aminoethyl)-cysteine (AEC) for 5 days.
[d] The mutants were generated in this study.
[e] Not determined.

The ability of the *metC* clones to express cystathionine β-lyase was tested by enzymatic assay. Crude extracts prepared from the *C. glutamicum* ASO19E12 cells harboring plasmid pSL173 were assayed. Cells harboring the plasmid showed approximately a 5-fold increase in the activity of cystathionine β-lyase compared to those harboring the empty vector pMT1 (Table 5), apparently due to the gene-dose effect. SDS-PAGE analysis of crude extracts revealed a putative cystathionine β-lyase band with approximate M_r of 41,000. Intensity of each putative cystathionine β-lyase band agreed with the complementation and enzymatic assay data (Table 5). As described above, a region of *metC* appeared to be nearly identical to the previously reported *aecD*. Since the *aecD* gene was isolated on the basis of its ability to confer resistance to *S*-(β-aminoethyl)-cysteine (AEC), a toxic lysine analogue, we tested the protein product of *metC* for the presence of the activity. As shown in Table 5, cells overexpressing cystathionine β-lyase showed increased resistance to AEC. The strain carrying a mutation in *metC* gene (see below) completely lost its ability to show a resistant phenotype to AEC.

Assay for *O*-acetylhmoserine sulphydrylase was performed as follows (Belfaiza, J., *et al.* (1998) *J. Bacteriol.* 180, 250-255; Ravanel, S., M. Droux, and R. Douce (1995) *Arch. Biochem. Biophys.* 316, 572-584; Foglino, M. (1995) *Microbiology* 141, 431-439).

Assay mixture of 0.1 ml contained 20 mM MOPS-NaOH (pH7.5), 10 mM *O*-acetylhomoserine, 2 mM Na_2S in 50 mM NaOH, and an appropriate amount of enzyme. Immediately after the addition of Na_2S which was added last, the reaction mixture was overlayed with 50 ul of mineral oil. After 30 minute incubation at 30°C, the reaction was stopped by boiling the mixture for 3 minutes. Homocysteine produced in the reaction was quantified as previously described (Yamagata, S. (1987) *Method Enzymol.* 143, 478-483.). Reaction mixture of 0.1 ml was taken and mixed with 0.1 ml of H_2O, 0.6 ml of saturated NaCl, 0.1 ml of 1.5 M Na_2CO_3 containing 67 mM KCN, and 0.1 ml of 2% nitroprusside. After 1 minute incubation at room temperature, optical density was measured at 520 nm. *Corynebacterium* cells harboring additional copies of the *metZ* gene, *e.g.*, a plasmid containing the *metZ* gene, exhibited significantly higher *metZ* enzyme activities than the same type of *Corynebacterium* cells without additional copies of the *metZ* gene.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. *et al.* (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) "Pores, Channels and Transporters", in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

Example 9: Analysis of Impact of Mutant Protein on the Production of the Desired Product

The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing the medium and/or the cellular component for increased production of the desired product (*i.e.*, an amino acid). Such analysis techniques are well known to one of ordinary skill in the art, and include spectroscopy, thin layer chromatography, staining

methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. *et al.*, (1987) "Applications of HPLC in Biochemistry" in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm *et al.* (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P.A. *et al.* (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (*e.g.*, sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

Example 10: Purification of the Desired Product from *C. glutamicum* Culture

Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum*

cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One of ordinary skill in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek *et al.* (1994) *Appl. Environ. Microbiol.* 60: 133-140; Malakhova *et al.* (1996) *Biotekhnologiya* 11: 27-32; and Schmidt *et al.* (1998) *Bioprocess Engineer.* 19: 67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. *et al.* (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

Example 11: Analysis of the Gene Sequences of the Invention

The comparison of sequences and determination of percent homology between two sequences are art-known techniques, and can be accomplished using a mathematical algorithm, such as the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci.* USA 87:2264-68, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci.* USA 90:5873-77. Such an algorithm is incorporated into the NBLAST and XBLAST

programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to MP nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to MP protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, one of ordinary skill in the art will know how to optimize the parameters of the program (*e.g.*, XBLAST and NBLAST) for the specific sequence being analyzed.

Another example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Meyers and Miller ((1988) *Comput. Appl. Biosci.* 4: 11-17). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art, and include ADVANCE and ADAM, described in Torelli and Robotti (1994) *Comput. Appl. Biosci.* 10:3-5; and FASTA, described in Pearson and Lipman (1988) *P.N.A.S.* 85:2444-8.

The percent homology between two amino acid sequences can also be accomplished using the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. The percent homology between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package, using standard parameters, such as a gap weight of 50 and a length weight of 3.

A comparative analysis of the gene sequences of the invention with those present in Genbank has been performed using techniques known in the art (see, *e.g.*, Bexevanis and Ouellette, eds. (1998) Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley and Sons: New York). The gene sequences of the invention were compared to genes present in Genbank in a three-step process. In a first step, a BLASTN analysis (*e.g.*, a local alignment analysis) was performed for each of the

sequences of the invention against the nucleotide sequences present in Genbank, and the top 500 hits were retained for further analysis. A subsequent FASTA search (*e.g.*, a combined local and global alignment analysis, in which limited regions of the sequences are aligned) was performed on these 500 hits. Each gene sequence of the invention was subsequently globally aligned to each of the top three FASTA hits, using the GAP program in the GCG software package (using standard parameters). In order to obtain correct results, the length of the sequences extracted from Genbank were adjusted to the length of the query sequences by methods well-known in the art. The results of this analysis are set forth in Table 4. The resulting data is identical to that which would have been obtained had a GAP (global) analysis alone been performed on each of the genes of the invention in comparison with each of the references in Genbank, but required significantly reduced computational time as compared to such a database-wide GAP (global) analysis. Sequences of the invention for which no alignments above the cutoff values were obtained are indicated on Table 4 by the absence of alignment information. It will further be understood by one of ordinary skill in the art that the GAP alignment homology percentages set forth in Table 4 under the heading "% homology (GAP)" are listed in the European numerical format, wherein a ',' represents a decimal point. For example, a value of "40,345" in this column represents "40.345%".

Example 12: Construction and Operation of DNA Microarrays

The sequences of the invention may additionally be used in the construction and application of DNA microarrays (the design, methodology, and uses of DNA arrays are well known in the art, and are described, for example, in Schena, M. *et al.* (1995) *Science* 270: 467-470; Wodicka, L. *et al.* (1997) *Nature Biotechnology* 15: 1359-1367; DeSaizieu, A. *et al.* (1998) *Nature Biotechnology* 16: 45-48; and DeRisi, J.L. *et al.* (1997) *Science* 278: 680-686).

DNA microarrays are solid or flexible supports consisting of nitrocellulose, nylon, glass, silicone, or other materials. Nucleic acid molecules may be attached to the surface in an ordered manner. After appropriate labeling, other nucleic acids or nucleic acid mixtures can be hybridized to the immobilized nucleic acid molecules, and the label may be used to monitor and measure the individual signal intensities of the hybridized molecules at defined regions. This methodology allows the simultaneous quantification of the relative or absolute amount of all or selected nucleic acids in the applied nucleic

acid sample or mixture. DNA microarrays, therefore, permit an analysis of the expression of multiple (as many as 6800 or more) nucleic acids in parallel (see, *e.g.*, Schena, M. (1996) *BioEssays* 18(5): 427-431).

The sequences of the invention may be used to design oligonucleotide primers which are able to amplify defined regions of one or more *C. glutamicum* genes by a nucleic acid amplification reaction such as the polymerase chain reaction. The choice and design of the 5' or 3' oligonucleotide primers or of appropriate linkers allows the covalent attachment of the resulting PCR products to the surface of a support medium described above (and also described, for example, Schena, M. *et al.* (1995) *Science* 270: 467-470).

Nucleic acid microarrays may also be constructed by *in situ* oligonucleotide synthesis as described by Wodicka, L. *et al.* (1997) *Nature Biotechnology* 15: 1359-1367. By photolithographic methods, precisely defined regions of the matrix are exposed to light. Protective groups which are photolabile are thereby activated and undergo nucleotide addition, whereas regions that are masked from light do not undergo any modification. Subsequent cycles of protection and light activation permit the synthesis of different oligonucleotides at defined positions. Small, defined regions of the genes of the invention may be synthesized on microarrays by solid phase oligonucleotide synthesis.

The nucleic acid molecules of the invention present in a sample or mixture of nucleotides may be hybridized to the microarrays. These nucleic acid molecules can be labeled according to standard methods. In brief, nucleic acid molecules (*e.g.*, mRNA molecules or DNA molecules) are labeled by the incorporation of isotopically or fluorescently labeled nucleotides, *e.g.*, during reverse transcription or DNA synthesis. Hybridization of labeled nucleic acids to microarrays is described (*e.g.*, in Schena, M. *et al.* (1995) *supra*; Wodicka, L. *et al.* (1997), *supra*; and DeSaizieu A. *et al.* (1998), *supra*). The detection and quantification of the hybridized molecule are tailored to the specific incorporated label. Radioactive labels can be detected, for example, as described in Schena, M. *et al.* (1995) *supra*) and fluorescent labels may be detected, for example, by the method of Shalon *et al.* (1996) *Genome Research* 6: 639-645).

The application of the sequences of the invention to DNA microarray technology, as described above, permits comparative analyses of different strains of *C. glutamicum* or other Corynebacteria. For example, studies of inter-strain variations

based on individual transcript profiles and the identification of genes that are important for specific and/or desired strain properties such as pathogenicity, productivity and stress tolerance are facilitated by nucleic acid array methodologies. Also, comparisons of the profile of expression of genes of the invention during the course of a fermentation reaction are possible using nucleic acid array technology.

Example 13: Analysis of the Dynamics of Cellular Protein Populations (Proteomics)

The genes, compositions, and methods of the invention may be applied to study the interactions and dynamics of populations of proteins, termed 'proteomics'. Protein populations of interest include, but are not limited to, the total protein population of *C. glutamicum* (*e.g.*, in comparison with the protein populations of other organisms), those proteins which are active under specific environmental or metabolic conditions (*e.g.*, during fermentation, at high or low temperature, or at high or low pH), or those proteins which are active during specific phases of growth and development.

Protein populations can be analyzed by various well-known techniques, such as gel electrophoresis. Cellular proteins may be obtained, for example, by lysis or extraction, and may be separated from one another using a variety of electrophoretic techniques. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins largely on the basis of their molecular weight. Isoelectric focusing polyacrylamide gel electrophoresis (IEF-PAGE) separates proteins by their isoelectric point (which reflects not only the amino acid sequence but also posttranslational modifications of the protein). Another, more preferred method of protein analysis is the consecutive combination of both IEF-PAGE and SDS-PAGE, known as 2-D-gel electrophoresis (described, for example, in Hermann *et al.* (1998) *Electrophoresis* 19: 3217-3221; Fountoulakis *et al.* (1998) *Electrophoresis* 19: 1193-1202; Langen *et al.* (1997) *Electrophoresis* 18: 1184-1192; Antelmann *et al.* (1997) *Electrophoresis* 18: 1451-1463). Other separation techniques may also be utilized for protein separation, such as capillary gel electrophoresis; such techniques are well known in the art.

Proteins separated by these methodologies can be visualized by standard techniques, such as by staining or labeling. Suitable stains are known in the art, and include Coomassie Brilliant Blue, silver stain, or fluorescent dyes such as Sypro Ruby (Molecular Probes). The inclusion of radioactively labeled amino acids or other protein

precursors (*e.g.*, ^{35}S-methionine, ^{35}S-cysteine, ^{14}C-labelled amino acids, ^{15}N-amino acids, $^{15}NO_3$ or $^{15}NH_4^+$ or ^{13}C-labelled amino acids) in the medium of *C. glutamicum* permits the labeling of proteins from these cells prior to their separation. Similarly, fluorescent labels may be employed. These labeled proteins can be extracted, isolated and separated according to the previously described techniques.

Proteins visualized by these techniques can be further analyzed by measuring the amount of dye or label used. The amount of a given protein can be determined quantitatively using, for example, optical methods and can be compared to the amount of other proteins in the same gel or in other gels. Comparisons of proteins on gels can be made, for example, by optical comparison, by spectroscopy, by image scanning and analysis of gels, or through the use of photographic films and screens. Such techniques are well-known in the art.

To determine the identity of any given protein, direct sequencing or other standard techniques may be employed. For example, N- and/or C-terminal amino acid sequencing (such as Edman degradation) may be used, as may mass spectrometry (in particular MALDI or ESI techniques (see, *e.g.*, Langen *et al.* (1997) *Electrophoresis* 18: 1184-1192)). The protein sequences provided herein can be used for the identification of *C. glutamicum* proteins by these techniques.

The information obtained by these methods can be used to compare patterns of protein presence, activity, or modification between different samples from various biological conditions (*e.g.*, different organisms, time points of fermentation, media conditions, or different biotopes, among others). Data obtained from such experiments alone, or in combination with other techniques, can be used for various applications, such as to compare the behavior of various organisms in a given (*e.g.*, metabolic) situation, to increase the productivity of strains which produce fine chemicals or to increase the efficiency of the production of fine chemicals.

Example 14: Cloning of Genes by Application of the Polymerase Chain Reaction (PCR)

Genes can be amplified using specific oligonucleotides comprising either nucleotide sequences homologous to sequences of *Corynebacterium glutamicum* or other strains as well as recognition sites of restriction enzymes well known in the art (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A*

Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Theses oligonucleotides can be used to amplify specific DNA-fragments containing parts of the chromosome of mentioned strains using DNA-polymerases such as *T. aquaticus* DNA-polymerase, *P. furiosus* DNA-polymerase, or *P. woesei* DNA-polymerase and dNTPs nucleotides in an appropriate buffer solution as described by the manufacturer.

Gene fragments such as coding sequences from RXA00657 including appropriate upstream and downstream regions not contained in the coding region of the mentioned gene can be amplified using the aforementioned technologies. Furthermore, these fragments can be purified from unincorporated oligonucleotides and nucleotides. DNA restriction enzymes can be used to produce protruding ends that can be used to ligate DNA fragments to vectors digested with complementary enzymes or compatible enzymes producing ends that can be used to ligate the DNA into the vectors mentioned in Sinskey *et al.*, U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (*e.g., lactofermentum*) (Yoshihama et al, *J. Bacteriol.* 162: 591-597 (1985); Katsumata *et al.*, *J. Bacteriol.* 159: 306-311 (1984); and Santamaria *et al.*, *J. Gen. Microbiol.* 130: 2237-2246 (1984). Oligonucleotides used as primers for the amplification of upstream DNA sequence, the coding region sequence and the downstream region of RXA00657 were as follows:

TCGGGTATCCGCGCTACACTTAGA (SEQ ID NO:121);
GGAAACCGGGGCATCGAAACTTA (SEQ ID NO:122).

Corynebacterium glutamicum chromosomal DNA with an amount of 200ng was used as a template in a 100μl reaction volume containing 2,5U Pfu Turbo-Polymerase™ (Stratagene™), and 200μM dNTP-nucleotides The PCR was performed on a PCR-Cycler™ (Perkin Elmer 2400™) using the following temperature/time protocol:

1 cycle: 94 °C: 2 min.;
20 cycle: 94°C : 1 min.;
52°C: 1 min, 72°C: 1.5 min.,
1 cycle: 72 °C: 5 min.

Primers were removed from the resulting amplified DNA fragment and the resulting fragment was cloned into the blunt EcoRV site of pBS KS (Stratagene™). The

fragment was excised by digestion with the restriction enzymes BamHI/XhoI and ligated into a BamHI SalI digested vector pB (SEQ ID NO.:125). The resulting vector is called pB RXA00657.

Resulting recombinant vectors can be analyzed using standard techniques described in *e.g.*, Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and can be transferred into *C. glutamicum* using aforementioned techniques.

A *Corynebacterium* strain (ATCC 13286) was treated for a transformation as described. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. *et al.* (1984) *J. Bacteriol.* 159306-311), electroporation (Liebl, E. *et al.* (1989) *FEMS Microbiol. Letters*, 53:399-303) and in cases where special vectors are used, also by conjugation (as described, *e.g.*, in Schäfer, A. *et al.* (1990) *J. Bacteriol.* 172:1663-1666). It is also possible to transfer the shuttle vectors for *C. glutamicum* to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) *J. Mol. Biol.* 166:1-19).

Transformation of a bacterial strain such as *Corynebacterium glutamicum* strain (ATCC 13286) was performed with a plasmid pB containing the aforementioned DNA regions of RXA00657 (SEQ ID NO.:6) and in another case with the vector pB (SEQ ID NO.:) carrying no additional insertion of nucleic acids.

The resulting strains were plated on and isolated from CM-Medium (10 g/l Glucose 2,5 g/l NaCl, 2,0 g/l Urea, 10 g/l Bacto Peptone (Difco/Becton Dicinson/Sparks USA™), 5 g/l yeast extract (Difco/Becton Dicinson/Sparks USA™), 5g/l meat extract (Difco/Becton Dicinson/Sparks USA™), 22g/l Agar (Difco/Becton Dickinson/Sparks USA™) and 15µg/ml kanamycin sulfate (Serva, Germany) with a adjusted with NaOH to pH of 6.8.

Strains isolated from the aforementioned agar medium were inoculated in 10 ml in a 100ml shake flask containing no baffles in liquid medium containing 100 g/l sucrose 50g/l $(NH4)_2SO_4$, 2,5 g/l NaCl, 2,0 g/l Urea, 10 g/l Bacto Peptone (Difco/Becton Dickinson/Sparks USA), 5 g/l yeast extract (Difco/Becton Dickinson/Sparks USA), 5g/l meat extract (Difco/Becton Dickinson/Sparks USA), and

25g/l CaCO3 (Riedel de Haen, Germany) . Medium was a adjusted with NaOH to pH of 6.8.

Strains were incubated at 30°C for 48h. Supernatants of incubations were prepared by centrifugation 20´at 12,000 rpm in an Eppendorf™ microcentrifuge. Liquid supernatants were diluted and subjected to amino acid analysis (Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein).

The results are shown in Table 6, below.

Results: Table 6:

Strain ATCC 13286	**Plasmid contained**	**pB**	**pB RXA00657**
lysin produced (g/l)		13.5	14.93
Selectivity (mol lysine/ mol consumed Saccharose)		0.235	0.25

Equivalents

Those of ordinary skill in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

TABLE 1: Included Genes

Lysine biosynthesis

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
5	**6**	**RXA00657**				AMINOACID BIOSYNTHESIS REGULATOR
7	8	RXA02229	GR00653	2793	3617	DIAMINOPIMELATE EPIMERASE (EC 5.1.1.7)
9	10	RXS02970				ACETYLORNITHINE AMINOTRANSFERASE (EC 2 6.1.11)
11	12	F RXA01009	GR00287	4714	5943	ACETYLORNITHINE AMINOTRANSFERASE (EC 2.6.1.11)
13	14	RXC02390				MEMBRANE SPANNING PROTEIN INVOLVED IN LYSINE METABOLISM
15	16	RXC01796				MEMBRANE ASSOCIATED PROTEIN INVOLVED IN LYSINE METABOLISM
17	18	RXC01207				CYTOSOLIC PROTEIN INVOLVED IN METABOLISM OF LYSINE AND THREONINE
19	20	RXC00657				TRANSCRIPTIONAL REGULATOR INVOLVED IN LYSINE METABOLISM
21	22	RXC00552				CYTOSOLIC PROTEIN INVOLVED IN LYSINE METABOLISM

Lysine biosynthesis

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
23	24	RXA00534	GR00137	4758	3496	ASPARTOKINASE ALPHA AND BETA SUBUNITS (EC 2.7.2.4)
25	26	RXA00533	GR00137	3469	2438	ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC 1.2.1.11)
27	28	RXA02843	GR00842	543	4	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLATE N-SUCCINYLTRANSFERASE (EC 2.3.1.117)
29	30	RXA02022	GR00613	2063	3169	SUCCINYL-DIAMINOPIMELATE DESUCCINYLASE (EC 3.5.1.18)
31	32	RXA00044	GR00007	3458	4393	DIHYDRODIPICOLINATE SYNTHASE (EC 4.2.1.52)
33	34	RXA00863	GR00236	896	1639	DIHYDRODIPICOLINATE REDUCTASE (EC 1 3.1.26)
35	36	RXA00864	GR00236	1694	2443	probable 2,3-dihydrodipicolinate N-C6-lyase (cyclizing) (EC 4.3.3.-) - Corynebacterium glutamicum
37	38	RXA02843	GR00842	543	4	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLATE N-SUCCINYLTRANSFERASE (EC 2.3.1.117)
39	40	RXN00355	VV0135	31980	30961	MESO-DIAMINOPIMELATE D-DEHYDROGENASE
41	42	F RXA00352	GR00068	861	4	MESO-DIAMINOPIMELATE D-DEHYDROGENASE (EC 1.4.1.16)
43	44	RXA00972	GR00274	3	1379	DIAMINOPIMELATE DECARBOXYLASE (EC 4.1.1.20)
45	46	RXA02653	GR00752	5237	7234	DIAMINOPIMELATE DECARBOXYLASE (EC 4.1.1.20)
47	48	RXA01393	GR00408	4249	3380	LYSINE EXPORT REGULATOR PROTEIN
49	50	RXA00241	GR00036	5443	6945	L-LYSINE TRANSPORT PROTEIN
51	52	RXA01394	GR00408	4320	5018	LYSINE EXPORTER PROTEIN
53	54	RXA00865	GR00236	2647	3549	DIHYDRODIPICOLINATE SYNTHASE (EC 4.2.1.52)

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
55	56	RXS02021				2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLATE N-SUCCINYLTRANSFERASE (EC 2.3.1.117)
57	58	RXS02157				ACETYLORNITHINE AMINOTRANSFERASE (EC 2.6.1.11)
59	60	RXC00733				ABC TRANSPORTER ATP-BINDING PROTEIN INVOLVED IN LYSINE METABOLISM
61	62	RXC00861				PROTEIN INVOLVED IN LYSINE METABOLISM
63	64	RXC00866				ZN-DEPENDENT HYDROLASE INVOLVED IN LYSINE METABOLISM
65	66	RXC02095				ABC TRANSPORTER ATP-BINDING PROTEIN INVOLVED IN LYSINE METABOLISM
67	68	RXC03185				PROTEIN INVOLVED IN LYSINE METABOLISM

Metabolism of methionine and S-adenosyl methionine

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
1	**2**	**metZ or met**				O-ACETYLHOMOSERINE SULFHYDRYLASE (EC 4.2.99.10)
3	**4**	**metC**				Cystathionine-γ-lyase
69	70	RXA00115	GR00017	5359	4313	HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.31)
71	72	RXN00403	VV0086	70041	68911	HOMOSERINE O-ACETYLTRANSFERASE
73	74	F RXA00403	GR00088	723	1832	HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.11)
75	76	RXS03158				CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
77	78	F RXA00254	GR00038	2404	1811	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
79	80	RXA02532	GR00726	3085	2039	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
81	82	RXS03159				CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
83	84	F RXA02768	GR00770	1919	2521	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
85	86	RXA00216	GR00032	16286	15297	5-methyltetrahydrofolate-homocysteine methyltransferase (methionine synthetase)
87	94	RXA02197	GR00645	4552	4025	5-METHYLTETRAHYDROFOLATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.13)
89	90	RXN02198	VV0302	9228	11726	5-METHYLTETRAHYDROFOLATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.13)
91	91	F RXA02198	GR00646	2483	6	5-METHYLTETRAHYDROFOLATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.13)
93	94	RXN03074	VV0042	2238	1741	S-ADENOSYLMETHIONINE:2-DEMETHYLMENAQUINONE METHYLTRANSFERASE (EC 2.1.-.-)

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
95	96	F RXA02906	GR10044	1142	645	S-ADENOSYLMETHIONINE:2-DEMETHYLMENAQUINONE METHYLTRANSFERASE (EC 2.1.-.-)
97	98	RXN00132	VV0124	3612	5045	ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1)
99	100	F RXA00132	GR00020	7728	7624	ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1)
101	102	F RXA01371	GR00398	2339	3634	ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1)
103	104	RXN02085				5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.14)
105	106	F RXA02085	GR00629	3496	5295	5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.14)
107	108	F RXA02086	GR00629	5252	5731	5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.14)
109	110	RXN02648				5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.14)
111	112	F RXA02648	GR00751	5254	4730	5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.14)
113	114	F RXA02658	GR00752	14764	15447	5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.14)
115	116	RXC02238				PROTEIN INVOLVED IN METABOLISM OF S-ADENOSYLMETHIONINE, PURINES AND PANTOTHENATE
117	118	RXC00128				EXPORTED PROTEIN INVOLVED IN METABOLISM OF PYRIDIMES AND ADENOSYLHOMOCYSTEINE

S-2adenosyl methionine (SAM) Biosynthesis

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
119	120	RXA02240	GR00654	7160	8380	S-ADENOSYLMETHIONINE SYNTHETASE (EC 2.5.1.6)

TABLE 2: GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
A09073	ppg	Phosphoenol pyruvate carboxylase	Bachmann, B. et al. "DNA fragment coding for phosphoenolpyruvat corboxylase, recombinant DNA carrying said fragment, strains carrying the recombinant DNA and method for producing L-aminino acids using said strains," Patent: EP 0358940-A 3 03/21/90
A45579, A45581, A45583, A45585 A45587		Threonine dehydratase	Moeckel, B. et al. "Production of L-isoleucine by means of recombinant micro-organisms with deregulated threonine dehydratase," Patent: WO 9519442-A 5 07/20/95
AB003132	murC; ftsQ; ftsZ		Kobayashi, M. et al. "Cloning, sequencing, and characterization of the ftsZ gene from coryneform bacteria," *Biochem. Biophys. Res. Commun.*, 236(2):383-388 (1997)
AB015023	murC; ftsQ		Wachi, M. et al. "A murC gene from Coryneform bacteria," *Appl Microbiol. Biotechnol.*, 51(2):223-228 (1999)
AB018530	dtsR		Kimura, E. et al. "Molecular cloning of a novel gene, dtsR, which rescues the detergent sensitivity of a mutant derived from *Brevibacterium lactofermentum*," *Biosci. Biotechnol. Biochem.*, 60(10):1565-1570 (1996)
AB018531	dtsR1; dtsR2		
AB020624	murI	D-glutamate racemase	
AB023377	tkt	transketolase	
AB024708	gltB; gltD	Glutamine 2-oxoglutarate aminotransferase large and small subunits	
AB025424	acn	aconitase	
AB027714	rep	Replication protein	
AB027715	rep; aad	Replication protein; aminoglycoside adenyltransferase	
AF005242	argC	N-acetylglutamate-5-semialdehyde dehydrogenase	
AF005635	glnA	Glutamine synthetase	
AF030405	hisF	cyclase	
AF030520	argG	Argininosuccinate synthetase	
AF031518	argF	Ornithine carbamolytransferase	

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF036932	aroD	3-dehydroquinate dehydratase	
AF038548	pyc	Pyruvate carboxylase	
AF038651	dciAE; apt; rel	Dipeptide-binding protein; adenine phosphoribosyltransferase; GTP pyrophosphokinase	Wehmeier, L. et al. "The role of the Corynebacterium glutamicum rel gene in (p)ppGpp metabolism," *Microbiology,* 144:1853-1862 (1998)
AF041436	argR	Arginine repressor	
AF045998	impA	Inositol monophosphate phosphatase	
AF048764	argH	Argininosuccinate lyase	
AF049897	argC; argJ; argB; argD; argF; argR; argG; argH	N-acetylglutamylphosphate reductase; ornithine acetyltransferase; N-acetylglutamate kinase; acetylornithine transminase; ornithine carbamoyltransferase; arginine repressor; argininosuccinate synthase; argininosuccinate lyase	
AF050109	inhA	Enoyl-acyl carrier protein reductase	
AF050166	hisG	ATP phosphoribosyltransferase	
AF051846	hisA	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazolecarboxamide isomerase	
AF052652	metA	Homoserine O-acetyltransferase	Park, S. et al. "Isolation and analysis of metA, a methionine biosynthetic gene encoding homoserine acetyltransferase in Corynebacterium glutamicum," *Mol. Cells.,* 8(3):286-294 (1998)
AF053071	aroB	Dehydroquinate synthetase	
AF060558	hisH	Glutamine amidotransferase	
AF086704	hisE	Phosphoribosyl-ATP-pyrophosphohydrolase	
AF114233	aroA	5-enolpyruvylshikimate 3-phosphate synthase	
AF116184	panD	L-aspartate-alpha-decarboxylase precursor	Dusch, N. et al. "Expression of the Corynebacterium glutamicum panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in Escherichia coli," *Appl. Environ. Microbiol.,* 65(4)1530-1539 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF124518	aroD; aroE	3-dehydroquinase; shikimate dehydrogenase	
AF124600	aroC; aroK; aroB; pepQ	Chorismate synthase; shikimate kinase; 3-dehydroquinate synthase; putative cytoplasmic peptidase	
AF145897	inhA		
AF145898	inhA		
AJ001436	ectP	Transport of ectoine, glycine betaine, proline	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes: Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
AJ004934	dapD	Tetrahydrodipicolinate succinylase (incomplete[i])	Wehrmann, A. et al. "Different modes of diaminopimelate synthesis and their role in cell wall integrity: A study with Corynebacterium glutamicum," *J. Bacteriol.*, 180(12):3159-3165 (1998)
AJ007732	ppc; secG; amt; ocd; soxA	Phosphoenolpyruvate-carboxylase; ?; high affinity ammonium uptake protein; putative ornithine-cyclodecarboxylase; sarcosine oxidase	
AJ010319	ftsY, glnB, glnD; srp; amtP	Involved in cell division; PII protein; uridylyltransferase (uridylyl-removing enzmye); signal recognition particle; low affinity ammonium uptake protein	Jakoby, M. et al. "Nitrogen regulation in Corynebacterium glutamicum; Isolation of genes involved in biochemical characterization of corresponding proteins," *FEMS Microbiol.*, 173(2):303-310 (1999)
AJ132968	cat	Chloramphenicol acetyl transferase	
AJ224946	mqo	L-malate: quinone oxidoreductase	Molenaar, D. et al. "Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from Corynebacterium glutamicum," *Eur. J. Biochem.*, 254(2):395-403 (1998)
AJ238250	ndh	NADH dehydrogenase	
AJ238703	porA	Porin	Lichtinger, T. et al. "Biochemical and biophysical characterization of the cell wall porin of Corynebacterium glutamicum: The channel is formed by a low molecular mass polypeptide," *Biochemistry,* 37(43):15024-15032 (1998)
D17429		Transposable element IS31831	Vertes et al."Isolation and characterization of IS31831, a transposable element from Corynebacterium glutamicum," *Mol. Microbiol.*, 11(4):739-746 (1994)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
D84102	odhA	2-oxoglutarate dehydrogenase	Usuda, Y. et al. "Molecular cloning of the Corynebacterium glutamicum (Brevibacterium lactofermentum AJ12036) odhA gene encoding a novel type of 2-oxoglutarate dehydrogenase," *Microbiology,* 142:3347-3354 (1996)
E01358	hdh; hk	Homoserine dehydrogenase; homoserine kinase	Katsumata, R. et al. "Production of L-thereonine and L-isoleucine," Patent: JP 1987232392-A 1 10/12/87
E01359		Upstream of the start codon of homoserine kinase gene	Katsumata, R. et al. "Production of L-thereonine and L-isoleucine," Patent: JP 1987232392-A 2 10/12/87
E01375		Tryptophan operon	
E01376	trpL; trpE	Leader peptide; anthranilate synthase	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E01377		Promoter and operator regions of tryptophan operon	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E03937		Biotin-synthase	Hatakeyama, K. et al. "DNA fragment containing gene capable of coding biotin synthetase and its utilization," Patent: JP 1992278088-A 1 10/02/92
E04040		Diamino pelargonic acid aminotransferase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04041		Desthiobiotinsynthetase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04307		Flavum aspartase	Kurusu, Y. et al. "Gene DNA coding aspartase and utilization thereof," Patent: JP 1993030977-A 1 02/09/93
E04376		Isocitric acid lyase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 03/09/93
E04377		Isocitric acid lyase N-terminal fragment	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 03/09/93
E04484		Prephenate dehydratase	Sotouchi, N. et al. "Production of L-phenylalanine by fermentation," Patent: JP 1993076352-A 2 03/30/93
E05108		Aspartokinase	Fugono, N. et al. "Gene DNA coding Aspartokinase and its use," Patent: JP 1993184366-A 1 07/27/93

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E05112		Dihydro-dipichorinate synthetase	Hatakeyama, K. et al. "Gene DNA coding dihydrodipicolinic acid synthetase and its use," Patent: JP 1993184371-A 1 07/27/93
E05776		Diaminopimelic acid dehydrogenase	Kobayashi, M. et al. "Gene DNA coding Diaminopimelic acid dehydrogenase and its use," Patent: JP 1993284970-A 1 11/02/93
E05779		Threonine synthase	Kohama, K. et al. "Gene DNA coding threonine synthase and its use," Patent: JP 1993284972-A 1 11/02/93
E06110		Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06111		Mutated Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06146		Acetohydroxy acid synthetase	Inui, M. et al. "Gene capable of coding Acetohydroxy acid synthetase and its use," Patent: JP 1993344893-A 1 12/27/93
E06825		Aspartokinase	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06826		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06827		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E07701	secY		Honno, N. et al. "Gene DNA participating in integration of membraneous protein to membrane," Patent: JP 1994169780-A 1 06/21/94
E08177		Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08178, E08179, E08180, E08181, E08182		Feedback inhibition-released Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08232		Acetohydroxy-acid isomeroreductase	Inui, M. et al. "Gene DNA coding acetohydroxy acid isomeroreductase," Patent: JP 1994277067-A 1 10/04/94
E08234	secE		Asai, Y. et al. "Gene DNA coding for translocation machinery of protein," Patent: JP 1994277073-A 1 10/04/94
E08643		FT aminotransferase and desthiobiotin synthetase promoter region	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E08646		Biotin synthetase	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95
E08649		Aspartase	Kohama, K. et al "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031478-A 1 02/03/95
E08900		Dihydrodipicolinate reductase	Madori, M. et al. "DNA fragment containing gene coding Dihydrodipicolinate acid reductase and utilization thereof," Patent: JP 1995075578-A 1 03/20/95
E08901		Diaminopimelic acid decarboxylase	Madori, M. et al. "DNA fragment containing gene coding Diaminopimelic acid decarboxylase and utilization thereof," Patent: JP 1995075579-A 1 03/20/95
E12594		Serine hydroxymethyltransferase	Hatakeyama, K. et al. "Production of L-trypophan," Patent: JP 1997028391-A 1 02/04/97
E12760, E12759, E12758		transposase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12764		Arginyl-tRNA synthetase; diaminopimelic acid decarboxylase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12767		Dihydrodipicolinic acid synthetase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12770		aspartokinase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12773		Dihydrodipicolinic acid reductase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E13655		Glucose-6-phosphate dehydrogenase	Hatakeyama, K. et al. "Glucose-6-phosphate dehydrogenase and DNA capable of coding the same," Patent: JP 1997224661-A 1 09/02/97
L01508	IlvA	Threonine dehydratase	Moeckel, B. et al. "Functional and structural analysis of the threonine dehydratase of Corynebacterium glutamicum," *J. Bacteriol.*, 174:8065-8072 (1992)
L07603	EC 4.2.1.15	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	Chen, C. et al. "The cloning and nucleotide sequence of Corynebacterium glutamicum 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene," *FEMS Microbiol. Lett.*, 107:223-230 (1993)
L09232	IlvB; ilvN; ilvC	Acetohydroxy acid synthase large subunit; Acetohydroxy acid synthase small subunit; Acetohydroxy acid isomeroreductase	Keilhauer, C. et al. "Isoleucine synthesis in Corynebacterium glutamicum: molecular analysis of the ilvB-ilvN-ilvC operon," *J. Bacteriol.*, 175(17):5595-5603 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
L18874	PtsM	Phosphoenolpyruvate sugar phosphotransferase	Fouet, A et al. "Bacillus subtilis sucrose-specific enzyme II of the phosphotransferase system: expression in Escherichia coli and homology to enzymes II from enteric bacteria," *PNAS USA*, 84(24):8773-8777 (1987); Lee, J.K. et al. "Nucleotide sequence of the gene encoding the Corynebacterium glutamicum mannose enzyme II and analyses of the deduced protein sequence," *FEMS Microbiol. Lett.*, 119(1-2):137-145 (1994)
L27123	aceB	Malate synthase	Lee, H-S. et al. "Molecular characterization of aceB, a gene encoding malate synthase in Corynebacterium glutamicum," *J. Microbiol. Biotechnol.*, 4(4):256-263 (1994)
L27126		Pyruvate kinase	Jetten, M. S. et al. "Structural and functional analysis of pyruvate kinase from Corynebacterium glutamicum," *Appl Environ Microbiol*, 60(7):2501-2507 (1994)
L28760	aceA	Isocitrate lyase	
L35906	dtxr	Diphtheria toxin repressor	Oguiza, J.A. et al. "Molecular cloning, DNA sequence analysis, and characterization of the Corynebacterium diphtheriae dtxR from Brevibacterium lactofermentum," *J. Bacteriol.*, 177(2):465-467 (1995)
M13774		Prephenate dehydratase	Follettie, M.T. et al. "Molecular cloning and nucleotide sequence of the Corynebacterium glutamicum pheA gene," *J. Bacteriol.*, 167:695-702 (1986)
M16175	5S rRNA		Park, Y-H. et al. "Phylogenetic analysis of the coryneform bacteria by 56 rRNA sequences," *J. Bacteriol.*, 169:1801-1806 (1987)
M16663	trpE	Anthranilate synthase, 5' end	Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene*, 52:191-200 (1987)
M16664	trpA	Tryptophan synthase, 3'end	Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene*, 52:191-200 (1987)
M25819		Phosphoenolpyruvate carboxylase	O'Regan, M. et al. "Cloning and nucleotide sequence of the Phosphoenolpyruvate carboxylase-coding gene of Corynebacterium glutamicum ATCC13032," *Gene*, 77(2):237-251 (1989)
M85106		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen. Microbiol.*, 138:1167-1175 (1992)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
M85107, M85108		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen. Microbiol.*, 138:1167-1175 (1992)
M89931	aecD; brnQ; yhbw	Beta C-S lyase; branched-chain amino acid uptake carrier; hypothetical protein yhbw	Rossol, I. et al. "The Corynebacterium glutamicum aecD gene encodes a C-S lyase with alpha, beta-elimination activity that degrades aminoethylcysteine," *J. Bacteriol.*, 174(9):2968-2977 (1992); Tauch, A. et al. "Isoleucine uptake in Corynebacterium glutamicum ATCC 13032 is directed by the brnQ gene product," *Arch. Microbiol.*, 169(4):303-312 (1998)
S59299	trp	Leader gene (promoter)	Herry, D.M. et al. "Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of Corynebacterium glutamicum: identification of a mutation in the trp leader sequence," *Appl. Environ. Microbiol.*, 59(3):791-799 (1993)
U11545	trpD	Anthranilate phosphoribosyltransferase	O'Gara, J.P. and Dunican, L.K. (1994) Complete nucleotide sequence of the Corynebacterium glutamicum ATCC 21850 tpD gene." Thesis, Microbiology Department, University College Galway, Ireland.
U13922	cglIM; cglIR; clgIIR	Putative type II 5-cytosoine methyltransferase; putative type II restriction endonuclease; putative type I or type III restriction endonuclease	Schafer, A. et al. "Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from Corynebacterium glutamicum ATCC 13032 and analysis of its role in intergeneric conjugation with Escherichia coli," *J. Bacteriol.*, 176(23):7309-7319 (1994); Schafer, A. et al. "The Corynebacterium glutamicum cglIM gene encoding a 5-cytosine in an McrBC-deficient Escherichia coli strain," *Gene*, 203(2):95-101 (1997)
U14965	recA		
U31224	ppx		Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)
U31225	proC	L-proline: NADP+ 5-oxidoreductase	Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)
U31230	obg; proB; unkdh	?;gamma glutamyl kinase;similar to D-isomer specific 2-hydroxyacid dehydrogenases	Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
U31281	bioB	Biotin synthase	Serebriiskii, I.G., "Two new members of the bio B superfamily: Cloning, sequencing and expression of bio B genes of Methylobacillus flagellatum and Corynebacterium glutamicum," *Gene,* 175:15-22 (1996)
U35023	thtR; accBC	Thiosulfate sulfurtransferase; acyl CoA carboxylase	Jager, W. et al. "A Corynebacterium glutamicum gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins," *Arch Microbiol*, 166(2);76-82 (1996)
U43535	cmr	Multidrug resistance protein	Jager, W. et al. "A Corynebacterium glutamicum gene conferring multidrug resistance in the heterologous host Escherichia coli," *J. Bacteriol.*, 179(7):2449-2451 (1997)
U43536	clpB	Heat shock ATP-binding protein	
U53587	aphA-3	3'5''-aminoglycoside phosphotransferase	
U89648		Corynebacterium glutamicum unidentified sequence involved in histidine biosynthesis, partial sequence	
X04960	trpA, trpB; trpC; trpD; trpE; trpG; trpL	Tryptophan operon	Matsui, K. et al. "Complete nucleotide and deduced amino acid sequences of the Brevibacterium lactofermentum tryptophan operon," *Nucleic Acids Res.*, 14(24):10113-10114 (1986)
X07563	lys A	DAP decarboxylase (meso-diaminopimelate decarboxylase, EC 4.1.1.20)	Yeh, P. et al. "Nucleic sequence of the lysA gene of Corynebacterium glutamicum and possible mechanisms for modulation of its expression," *Mol. Gen. Genet.*, 212(1):112-119 (1988)
X14234	EC 4.1.1.31	Phosphoenolpyruvate carboxylase	Eikmanns, B.J. et al. "The Phosphoenolpyruvate carboxylase gene of Corynebacterium glutamicum: Molecular cloning, nucleotide sequence, and expression," *Mol Gen Genet*, 218(2):330-339 (1989); Lepiniec, L. et al. "Sorghum Phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution," *Plant. Mol. Biol.*, 21 (3):487-502 (1993)
X17313	fda	Fructose-bisphosphate aldolase	Von der Osten, C.H. et al. "Molecular cloning, nucleotide sequence and fine-structural analysis of the Corynebacterium glutamicum fda gene: structural comparison of C. glutamicum fructose-1, 6-biphosphate aldolase to class I and class II aldolases," *Mol Microbiol*,
X53993	dapA	L-2, 3-dihydrodipicolinate synthetase (EC 4.2.1.52)	Bonnassie, S. et al. "Nucleic sequence of the dapA gene from Corynebacterium glutamicum," *Nucleic Acids Res.*, 18(21):6421 (1990)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X54223		AttB-related site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum , and the attP site of lambdacorynephage," *FEMS. Microbiol, Lett.*, 66:299-302 (1990)
X54740	argS; lysA	Arginyl-tRNA synthetase; Diaminopimelate decarboxylase	Marcel, T. et al. "Nucleotide sequence and organization of the upstream region of the Corynebacterium glutamicum lysA gene," *Mol. Microbiol.*, 4(11):1819-1830 (1990)
X55994	trpL; trpE	Putative leader peptide; anthranilate synthase component 1	Heery, D.M. et al. "Nucleotide sequence of the Corynebacterium glutamicum trpE gene," *Nucleic Acids Res.*, 18(23):7138 (1990)
X56037	thrC	Threonine synthase	Han, K.S. et al. "The molecular structure of the Corynebacterium glutamicum threonine synthase gene," *Mol. Microbiol.*, 4(10):1693-1702 (1990)
X56075	attB-related site	Attachment site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum , and the attP site of lambdacorynephage," *FEMS. Microbiol, Lett.*, 66:299-302 (1990)
X57226	lysC-alpha; lysC-beta; asd	Aspartokinase-alpha subunit; Aspartokinase-beta subunit; aspartate beta semialdehyde dehydrogenase	Kalinowski, J. et al. "Genetic and biochemical analysis of the Aspartokinase from Corynebacterium glutamicum," *Mol. Microbiol*, 5(5):1197-1204 (1991); Kalinowski, J. et al. "Aspartokinase genes lysC alpha and lysC beta overlap and are adjacent to the aspertate beta-semialdehyde dehydrogenase gene asd in Corynebacterium glutamicum," *Mol. Gen. Genet.*, 224(3):317-324 (1990)
X59403	gap;pgk; tpi	Glyceraldehyde-3-phosphate; phosphoglycerate kinase; triosephosphate isomerase	Eikmanns, B.J. "Identification, sequence analysis, and expression of a Corynebacterium glutamicum gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomeras," *J. Bacteriol.*, 174(19):6076-6086 (1992)
X59404	gdh	Glutamate dehydrogenase	Bormann, E.R. et al. "Molecular analysis of the Corynebacterium glutamicum gdh gene encoding glutamate dehydrogenase," *Mol. Microbiol.*, 6(3):317-326 (1992)
X60312	lysI	L-lysine permease	Seep-Feldhaus, A.H. et al. "Molecular analysis of the Corynebacterium glutamicum lysI gene involved in lysine uptake," *Mol. Microbiol.*, 5(12):2995-3005 (1991)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X66078	cop1	Ps1 protein	Joliff, G. et al. "Cloning and nucleotide sequence of the csp1 gene encoding PS1, one of the two major secreted proteins of Corynebacterium glutamicum: The deduced N-terminal region of PS1 is similar to the Mycobacterium antigen 85 complex," *Mol. Microbiol.*, 6(16):2349-2362 (1992)
X66112	glt	Citrate synthase	Eikmanns, B.J. et al. "Cloning sequence, expression and transcriptional analysis of the Corynebacterium glutamicum gltA gene encoding citrate synthase," *Microbiol.*, 140:1817-1828 (1994)
X67737	dapB	Dihydrodipicolinate reductase	
X69103	csp2	Surface layer protein PS2	Peyret, J.L. et al. "Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in Corynebacterium glutamicum," *Mol. Microbiol.*, 9(1):97-109 (1993)
X69104		IS3 related insertion element	Bonamy, C. et al. "Identification of IS1206, a Corynebacterium glutamicum IS3-related insertion sequence and phylogenetic analysis," *Mol. Microbiol.*, 14(3):571-581 (1994)
X70959	leuA	Isopropylmalate synthase	Patek, M. et al. "Leucine synthesis in Corynebacterium glutamicum: enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis," *Appl. Environ. Microbiol.*, 60(1):133-140 (1994)
X71489	icd	Isocitrate dehydrogenase (NADP+)	Eikmanns, B.J. et al. "Cloning sequence analysis, expression, and inactivation of the Corynebacterium glutamicum icd gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme," *J. Bacteriol.*, 177(3):774-782 (1995)
X72855	GDHA	Glutamate dehydrogenase (NADP+)	
X75083, X70584	mtrA	5-methyltryptophan resistance	Heery, D.M. et al. "A sequence from a tryptophan-hyperproducing strain of Corynebacterium glutamicum encoding resistance to 5-methyltryptophan," *Biochem. Biophys. Res. Commun.*, 201(3):1255-1262 (1994)
X75085	recA		Fitzpatrick, R. et al. "Construction and characterization of recA mutant strains of Corynebacterium glutamicum and Brevibacterium lactofermentum," *Appl. Microbiol. Biotechnol.*, 42(4):575-580 (1994)
X75504	aceA; thiX	Partial Isocitrate lyase; ?	Reinscheid, D.J. et al. "Characterization of the isocitrate lyase gene from Corynebacterium glutamicum and biochemical analysis of the enzyme," *J. Bacteriol.*, 176(12):3474-3483 (1994)
X76875		ATPase beta-subunit	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit

			genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)
GenBank™ Accession No.	**Gene Name**	**Gene Function**	**Reference**
X77034	tuf	Elongation factor Tu	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)
X77384	recA		Billman-Jacobe, H. "Nucleotide sequence of a recA gene from Corynebacterium glutamicum," *DNA Seq.*, 4(6):403-404 (1994)
X78491	aceB	Malate synthase	Reinscheid, D.J. et al. "Malate synthase from Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase: sequence analysis," *Microbiology*, 140:3099-3108 (1994)
X80629	16S rDNA	16S ribosomal RNA	Rainey, F.A. et al. "Phylogenetic analysis of the genera Rhodococcus and Norcardia and evidence for the evolutionary origin of the genus Norcardia from within the radiation of Rhodococcus species," *Microbiol.*, 141:523-528 (1995)
X81191	gluA; gluB; gluC; gluD	Glutamate uptake system	Kronemeyer, W. et al. "Structure of the gluABCD cluster encoding the glutamate uptake system of Corynebacterium glutamicum," *J. Bacteriol.*, 177(5):1152-1158 (1995)
X81379	dapE	Succinyldiaminopimelate desuccinylase	Wehrmann, A. et al. "Analysis of different DNA fragments of Corynebacterium glutamicum complementing dapE of Escherichia coli," *Microbiology*, 40:3349-56 (1994)
X82061	16S rDNA	16S ribosomal RNA	Ruimy, R. et al. "Phylogeny of the genus Corynebacterium deduced from analyses of small-subunit ribosomal DNA sequences," *Int. J. Syst. Bacteriol.*, 45(4):740-746 (1995)
X82928	asd; lysC	Aspartate-semialdehyde dehydrogenase; ?	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J. Bacteriol.*, 177(24):7255-7260 (1995)
X82929	proA	Gamma-glutamyl phosphate reductase	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J. Bacteriol.*, 177(24):7255-7260 (1995)
X84257	16S rDNA	16S ribosomal RNA	Pascual, C. et al. "Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences," *Int. J. Syst. Bacteriol.*, 45(4):724-728 (1995)
X85965	aroP; dapE	Aromatic amino acid permease; ?	Wehrmann et al. "Functional analysis of sequences adjacent to dapE of C. glutamicum proline reveals the presence of aroP, which encodes the aromatic amino acid transporter," *J. Bacteriol.*, 177(20):5991-5993 (1995)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X86157	argB; argC; argD; argF; argJ	Acetylglutamate kinase; N-acetyl-gamma-glutamyl-phosphate reductase; acetylornithine aminotransferase; ornithine carbamoyltransferase; glutamate N-acetyltransferase	Sakanyan, V. et al. "Genes and enzymes of the acetyl cycle of arginine biosynthesis in Corynebacterium glutamicum: enzyme evolution in the early steps of the arginine pathway," *Microbiology,* 142:99-108 (1996)
X89084	pta; ackA	Phosphate acetyltransferase; acetate kinase	Reinscheid, D.J. et al. "Cloning, sequence analysis, expression and inactivation of the Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase and acetate kinase," *Microbiology,* 145:503-513 (1999)
X89850	attB	Attachment site	Le Marrec, C. et al. "Genetic characterization of site-specific integration functions of phi AAU2 infecting "Arthrobacter aureus C70," *J. Bacteriol.,* 178(7):1996-2004 (1996)
X90356		Promoter fragment F1	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90357		Promoter fragment F2	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90358		Promoter fragment F10	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90359		Promoter fragment F13	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90360		Promoter fragment F22	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90361		Promoter fragment F34	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)
X90362		Promoter fragment F37	Patek, M. et al. "Promoters from C. glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology,* 142:1297-1309 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X90363		Promoter fragment F45	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90364		Promoter fragment F64	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90365		Promoter fragment F75	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90366		Promoter fragment PF101	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90367		Promoter fragment PF104	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90368		Promoter fragment PF109	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X93513	amt	Ammonium transport system	Siewe, R.M. et al. "Functional and genetic characterization of the (methyl) ammonium uptake carrier of Corynebacterium glutamicum," *J. Biol. Chem.*, 271(10):5398-5403 (1996)
X93514	betP	Glycine betaine transport system	Peter, H. et al. "Isolation, characterization, and expression of the Corynebacterium glutamicum betP gene, encoding the transport system for the compatible solute glycine betaine," *J. Bacteriol.*, 178(17):5229-5234 (1996)
X95649	orf4		Patek, M. et al. "Identification and transcriptional analysis of the dapB-ORF2-dapA-ORF4 operon of Corynebacterium glutamicum, encoding two enzymes involved in L-lysine synthesis," *Biotechnol. Lett.*, 19:1113-1117 (1997)
X96471	lysE; lysG	Lysine exporter protein; Lysine export regulator protein	Vrljic, M. et al. "A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum," *Mol. Microbiol.*, 22(5):815-826 (1996)
X96580	panB; panC; xylB	3-methyl-2-oxobutanoate hydroxymethyltransferase; pantoate-beta-	Sahm, H. et al. "D-pantothenate synthesis in Corynebacterium glutamicum and use of panBC and genes encoding L-valine synthesis for D-pantothenate

		alanine ligase; xylulokinase	overproduction," *Appl. Environ. Microbiol.*, 65(5):1973-1979 (1999)
GenBank™ Accession No.	**Gene Name**	**Gene Function**	**Reference**
X96962		Insertion sequence IS1207 and transposase	
X99289		Elongation factor P	Ramos, A. et al. "Cloning, sequencing and expression of the gene encoding elongation factor P in the amino-acid producer Brevibacterium lactofermentum (Corynebacterium glutamicum ATCC 13869)," *Gene,* 198:217-222 (1997)
Y00140	thrB	Homoserine kinase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine kinase (thrB) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(9):3922 (1987)
Y00151	ddh	Meso-diaminopimelate D-dehydrogenase (EC 1.4.1.16)	Ishino, S. et al. "Nucleotide sequence of the meso-diaminopimelate D-dehydrogenase gene from Corynebacterium glutamicum," *Nucleic Acids Res.*, 15(9):3917 (1987)
Y00476	thrA	Homoserine dehydrogenase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine dehydrogenase (thrA) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(24):10598 (1987)
Y00546	hom; thrB	Homoserine dehydrogenase; homoserine kinase	Peoples, O.P. et al. "Nucleotide sequence and fine structural analysis of the Corynebacterium glutamicum hom-thrB operon," *Mol. Microbiol.*, 2(1):63-72 (1988)
Y08964	murC; ftsQ/divD; ftsZ	UPD-N-acetylmuramate-alanine ligase; division initiation protein or cell division protein; cell division protein	Honrubia, M.P. et al. "Identification, characterization, and chromosomal organization of the ftsZ gene from Brevibacterium lactofermentum," *Mol. Gen. Genet.*, 259(1):97-104 (1998)
Y09163	putP	High affinity proline transport system	Peter, H. et al. "Isolation of the putP gene of Corynebacterium glutamicumproline and characterization of a low-affinity uptake system for compatible solutes," *Arch. Microbiol.*, 168(2):143-151 (1997)
Y09548	pyc	Pyruvate carboxylase	Peters-Wendisch, P.G. et al. "Pyruvate carboxylase from Corynebacterium glutamicum: characterization, expression and inactivation of the pyc gene," *Microbiology*, 144:915-927 (1998)
Y09578	leuB	3-isopropylmalate dehydrogenase	Patek, M. et al. "Analysis of the leuB gene from Corynebacterium glutamicum," *Appl Microbiol Biotechnol.*, 50(1):42-47 (1998)
Y12472		Attachment site bacteriophage Phi-16	Moreau, S. et al. "Site-specific integration of corynephage Phi-16: The construction of an integration vector," *Microbiol.*, 145:539-548 (1999)
Y12537	proP	Proline/ectoine uptake system protein	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes: Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
Y13221	glnA	Glutamine synthetase I	Jakoby, M. et al. "Isolation of Corynebacterium glutamicum glnA gene encoding glutamine synthetase I," *FEMS Microbiol. Lett.*, 154(1):81-88 (1997)
Y16642	lpd	Dihydrolipoamide dehydrogenase	
Y18059		Attachment site Corynephage 304L	Moreau, S. et al. "Analysis of the integration functions of φ304L: An integrase module among corynephages," *Virology*, 255(1):150-159 (1999)
Z21501	argS; lysA	Arginyl-tRNA synthetase; diaminopimelate decarboxylase (partial)	Oguiza, J.A. et al. "A gene encoding arginyl-tRNA synthetase is located in the upstream region of the lysA gene in Brevibacterium lactofermentum: Regulation of argS-lysA cluster expression by arginine," *J. Bacteriol.*,175(22):7356-7362 (1993)
Z21502	dapA; dapB	Dihydrodipicolinate synthase; dihydrodipicolinate reductase	Pisabarro, A. et al. "A cluster of three genes (dapA, orf2, and dapB) of Brevibacterium lactofermentum encodes dihydrodipicolinate reductase, and a third polypeptide of unknown function," *J. Bacteriol.*, 175(9):2743-2749 (1993)
Z29563	thrC	Threonine synthase	Malumbres, M. et al. "Analysis and expression of the thrC gene of the encoded threonine synthase," *Appl Environ Microbiol*, 60(7)2209-2219 (1994)
Z46753	16S rDNA	Gene for 16S ribosomal RNA	
Z49822	sigA	SigA sigma factor	Oguiza, J.A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J. Bacteriol.*, 178(2):550-553 (1996)
Z49823	galE; dtxR	Catalytic activity UDP-galactose 4-epimerase; diphtheria toxin regulatory protein	Oguiza, J.A. et al "The galE gene encoding the UDP-galactose 4-epimerase of Brevibacterium lactofermentum is coupled transcriptionally to the dmdR gene," *Gene*, 177:103-107 (1996)
Z49824	orf1; sigB	?; SigB sigma factor	Oguiza, J.A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J. Bacteriol.*, 178(2):550-553 (1996)
Z66534		Transposase	Correia, A. et al. "Cloning and characterization of an IS-like element present in the genome of Brevibacterium lactofermentum ATCC 13869," *Gene*, 170(1):91-94 (1996)

[1] A sequence for this gene was published in the indicated reference. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

TABLE 3: Corynebacterium and Brevibacterium Strains Which May be Used in the Practice of the Invention

Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ	
Brevibacterium	ammoniagenes	21054								
Brevibacterium	ammoniagenes	19350								
Brevibacterium	ammoniagenes	19351								
Brevibacterium	ammoniagenes	19352								
Brevibacterium	ammoniagenes	19353								
Brevibacterium	ammoniagenes	19354								
Brevibacterium	ammoniagenes	19355								
Brevibacterium	ammoniagenes	19356								
Brevibacterium	ammoniagenes	21055								
Brevibacterium	ammoniagenes	21077								
Brevibacterium	ammoniagenes	21553								
Brevibacterium	ammoniagenes	21580								
Brevibacterium	ammoniagenes	39101								
Brevibacterium	butanicum	21196								
Brevibacterium	divaricatum	21792	P928							
Brevibacterium	flavum	21474								
Brevibacterium	flavum	21129								
Brevibacterium	flavum	21518								
Brevibacterium	flavum			B11474						
Brevibacterium	flavum			B11472						
Brevibacterium	flavum	21127								
Brevibacterium	flavum	21128								
Brevibacterium	flavum	21427								
Brevibacterium	flavum	21475								
Brevibacterium	flavum	21517								
Brevibacterium	flavum	21528								
Brevibacterium	flavum	21529								
Brevibacterium	flavum			B11477						
Brevibacterium	flavum			B11478						
Brevibacterium	flavum	21127								
Brevibacterium	flavum			B11474						
Brevibacterium	healii	15527								
Brevibacterium	ketoglutamicum	21004								
Brevibacterium	ketoglutamicum	21089								
Brevibacterium	ketosoreductum	21914								
Brevibacterium	lactofermentum				70					
Brevibacterium	lactofermentum				74					
Brevibacterium	lactofermentum				77					
Brevibacterium	lactofermentum	21798								
Brevibacterium	lactofermentum	21799								
Brevibacterium	lactofermentum	21800								
Brevibacterium	lactofermentum	21801								
Brevibacterium	lactofermentum			B11470						
Brevibacterium	lactofermentum			B11471						

Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ	
Brevibacterium	lactofermentum	21086								
Brevibacterium	lactofermentum	21420								
Brevibacterium	lactofermentum	21086								
Brevibacterium	lactofermentum	31269								
Brevibacterium	linens	9174								
Brevibacterium	linens	19391								
Brevibacterium	linens	8377								
Brevibacterium	paraffinolyticum					11160				
Brevibacterium	spec.						717.73			
Brevibacterium	spec.						717.73			
Brevibacterium	spec.	14604								
Brevibacterium	spec.	21860								
Brevibacterium	spec.	21864								
Brevibacterium	spec.	21865								
Brevibacterium	spec.	21866								
Brevibacterium	spec.	19240								
Corynebacterium	acetoacidophilum	21476								
Corynebacterium	acetoacidophilum	13870								
Corynebacterium	acetoglutamicum			B11473						
Corynebacterium	acetoglutamicum			B11475						
Corynebacterium	acetoglutamicum	15806								
Corynebacterium	acetoglutamicum	21491								
Corynebacterium	acetoglutamicum	31270								
Corynebacterium	acetophilum			B3671						
Corynebacterium	ammoniagenes	6872						2399		
Corynebacterium	ammoniagenes	15511								
Corynebacterium	fujiokense	21496								
Corynebacterium	glutamicum	14067								
Corynebacterium	glutamicum	39137								
Corynebacterium	glutamicum	21254								
Corynebacterium	glutamicum	21255								
Corynebacterium	glutamicum	31830								
Corynebacterium	glutamicum	13032								
Corynebacterium	glutamicum	14305								
Corynebacterium	glutamicum	15455								
Corynebacterium	glutamicum	13058								
Corynebacterium	glutamicum	13059								
Corynebacterium	glutamicum	13060								
Corynebacterium	glutamicum	21492								
Corynebacterium	glutamicum	21513								
Corynebacterium	glutamicum	21526								
Corynebacterium	glutamicum	21543								
Corynebacterium	glutamicum	13287								
Corynebacterium	glutamicum	21851								
Corynebacterium	glutamicum	21253								
Corynebacterium	glutamicum	21514								
Corynebacterium	glutamicum	21516								
Corynebacterium	glutamicum	21299								

Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ	
Corynebacterium	glutamicum	21300								
Corynebacterium	glutamicum	39684								
Corynebacterium	glutamicum	21488								
Corynebacterium	glutamicum	21649								
Corynebacterium	glutamicum	21650								
Corynebacterium	glutamicum	19223								
Corynebacterium	glutamicum	13869								
Corynebacterium	glutamicum	21157								
Corynebacterium	glutamicum	21158								
Corynebacterium	glutamicum	21159								
Corynebacterium	glutamicum	21355								
Corynebacterium	glutamicum	31808								
Corynebacterium	glutamicum	21674								
Corynebacterium	glutamicum	21562								
Corynebacterium	glutamicum	21563								
Corynebacterium	glutamicum	21564								
Corynebacterium	glutamicum	21565								
Corynebacterium	glutamicum	21566								
Corynebacterium	glutamicum	21567								
Corynebacterium	glutamicum	21568								
Corynebacterium	glutamicum	21569								
Corynebacterium	glutamicum	21570								
Corynebacterium	glutamicum	21571								
Corynebacterium	glutamicum	21572								
Corynebacterium	glutamicum	21573								
Corynebacterium	glutamicum	21579								
Corynebacterium	glutamicum	19049								
Corynebacterium	glutamicum	19050								
Corynebacterium	glutamicum	19051								
Corynebacterium	glutamicum	19052								
Corynebacterium	glutamicum	19053								
Corynebacterium	glutamicum	19054								
Corynebacterium	glutamicum	19055								
Corynebacterium	glutamicum	19056								
Corynebacterium	glutamicum	19057								
Corynebacterium	glutamicum	19058								
Corynebacterium	glutamicum	19059								
Corynebacterium	glutamicum	19060								
Corynebacterium	glutamicum	19185								
Corynebacterium	glutamicum	13286								
Corynebacterium	glutamicum	21515								
Corynebacterium	glutamicum	21527								
Corynebacterium	glutamicum	21544								
Corynebacterium	glutamicum	21492								
Corynebacterium	glutamicum			B8183						
Corynebacterium	glutamicum			B8182						
Corynebacterium	glutamicum			B12416						
Corynebacterium	glutamicum			B12417						

Genus	**species**	**ATCC**	**FERM**	**NRRL**	**CECT**	**NCIMB**	**CBS**	**NCTC**	**DSMZ**	**Other origin**
Corynebacterium	glutamicum			B12418						
Corynebacterium	glutamicum			B11476						
Corynebacterium	glutamicum	21608								
Corynebacterium	lilium		P973							
Corynebacterium	nitrilophilus	21419				11594				
Corynebacterium	spec.		P4445							
Corynebacterium	spec		P4446							
Corynebacterium	spec.	31088								
Corynebacterium	spec.	31089								
Corynebacterium	spec	31090								
Corynebacterium	spec	31090								
Corynebacterium	spec	31090								
Corynebacterium	spec.	15954							20145	
Corynebacterium	spec.	21857								
Corynebacterium	spec.	21862								
Corynebacterium	spec.	21863								
Corynebacterium	Glutamicum*									ASO19
Corynebacterium	Glutamicum**									ASO19 E12
Corynebacterium	Glutamicum***									HL457
Corynebacterium	Glutamicum****									HL459

ATCC: American Type Culture Collection, Rockville, MD, USA

FERM: Fermentation Research Institute, Chiba, Japan

NRRL: ARS Culture Collection, Northern Regional Research Laboratory, Peoria, IL, USA

CECT: Coleccion Espanola de Cultivos Tipo, Valencia, Spain

NCIMB: National Collection of Industrial and Marine Bacteria Ltd., Aberdeen, UK

CBS: Centraalbureau voor Schimmelcultures, Baarn, NL

NCTC: National Collection of Type Cultures, London, UK

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany

For reference see Sugawara, H. et al. (1993) World directory of collections of cultures of microorganisms: Bacteria, fungi and yeasts (4th edn), World federation for culture collections world data center on microorganisms, Saimata, Japen.

* Spontaneous rifampin-resistant mutant of *C. glutamicum* ATCC13059[d] Yoshihama *et al.*, 1985

** Restriction-deficient variant of ASO19 Follettie *et al.*, 1993

*** *metC*-disrupted mutant of ASO19E12 This study

**** *metC*-disrupted mutant of ASO19E12 This study

TABLE 4: ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa00657	906	GB_BA1:AF064700	3481	AF064700	Rhodococcus sp. NO1-1 CprS and CprR genes, complete cds.	Rhodococcus sp	40.265	15-Jul-98
metz	1314	GB_BA2:MTV016	53662	AL021841	Mycobacterium tuberculosis H37Rv complete genome; segment 143/162.	Mycobacterium tuberculosis	61.278	23-Jun-99
metc	978	GB_BA2:CORCSLYS	2821	M89931	Corynebacterium glutamicum beta C-S lyase (aecD) and branched-chain amino acid upta	Corynebacterium glutamicum	99.591	04-JUN-1998
rxa00023	3579	GB_EST33:AI776129	483	AI776129	EST257217 tomato resistant, Cornell Lycopersicon esculentum cDNA clone cLER17D3, mRNA sequence.	Lycopersicon esculentum	40,956	29-Jun-99
		GB_EST33:AI776129	483	AI776129	EST257217 tomato resistant, Cornell Lycopersicon esculentum cDNA clone cLER17D3, mRNA sequence.	Lycopersicon esculentum	40,956	29-Jun-99
rxa00044	1059	EM_PAT:E11760	6911	E11760	Base sequence of sucrase gene.	Corynebacterium glutamicum	42,979	08-OCT-1997 (Rel. 52, Created)
		GB_PAT:I26124	6911	I26124	Sequence 4 from patent US 5556776.	Unknown.	42,979	07-OCT-1996
		GB_BA2:ECOUW89	176195	U00006	E. coli chromosomal region from 89.2 to 92.8 minutes.	Escherichia coli	39,097	17-DEC-1993
rxa00064	1401	GB_PAT:E16763	2517	E16763	gDNA encoding aspartate transferase (AAT).	Corynebacterium glutamicum	95,429	28-Jul-99
		GB_HTG2:AC007892	134257	AC007892	Drosophila melanogaster chromosome 3 clone BACR02O03 (D797) RPCI-98 02.O.3 map 99B-99B strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 113 unordered pieces.	Drosophila melanogaster	31,111	2-Aug-99
		GB_HTG2:AC007892	134257	AC007892	Drosophila melanogaster chromosome 3 clone BACR02O03 (D797) RPCI-98 02.O.3 map 99B-99B strain y; cn bw sp, *** SEQUENCING IN PROGRESS***, 113 unordered pieces.	Drosophila melanogaster	31,111	2-Aug-99
rxa00072								
rxa00105	798	GB_BA1:MTV002	56414	AL008967	Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.	Mycobacterium tuberculosis	37,753	17-Jun-98
		GB_BA1:ECU29581	71128	U29581	Escherichia coli K-12 genome; approximately 63 to 64 minutes.	Escherichia coli	35,669	14-Jan-97
		GB_BA2:AE000366	10405	AE000366	Escherichia coli K-12 MG1655 section 256 of 400 of the complete genome.	Escherichia coli	35,669	12-Nov-98
rxa00106	579	GB_EST15:AA494237	367	AA494237	ng83f04.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:941407 similar to SW:DYR_LACCA P00381 DIHYDROFOLATE REDUCTASE ;, mRNA sequence.	Homo sapiens	42,896	20-Aug-97
		GB_BA2:AF161327	2021	AF161327	Corynebacterium diphtheriae histidine kinase ChrS (chrS) and response regulator ChrA (chrA) genes, complete cds.	Corynebacterium diphtheriae	40,210	9-Sep-99
		GB_PAT:AR041189	654	AR041189	Sequence 4 from patent US 5811286.	Unknown.	41,176	29-Sep-99
rxa00115	1170	GB_PR4:AC007110	148336	AC007110	Homo sapiens chromosome 17, clone hRPK.472_J_18, complete sequence.	Homo sapiens	36,783	30-MAR-1999
		GB_HTG3:AC008537	170030	AC008537	Homo sapiens chromosome 19 clone CIT-HSPC_490E21, *** SEQUENCING IN PROGRESS ***, 93 unordered pieces.	Homo sapiens	40,296	2-Sep-99
		GB_HTG3:AC008537	170030	AC008537	Homo sapiens chromosome 19 clone CIT-HSPC_490E21, *** SEQUENCING IN PROGRESS ***, 93 unordered pieces.	Homo sapiens	40,296	2-Sep-99

TABLE 4: ALIGNMENT RESULTS

rxa00116	1284	GB_BA2:AF062345	16458	AF062345	Caulobacter crescentus Sst1 (sst1), S-layer protein subunit (rsaA), ABC transporter (rsaD), membrane forming unit (rsaE), putative GDP-mannose-4,6-dehydratase (lpsA), putative acetyltransferase (lpsB), putative perosamine synthetase (lpsC), putative mannosyltransferase (lpsD), putative mannosyltransferase (lpsE), outer membrane protein (rsaF), and putative perosamine transferase (lpsE) genes, complete cds.	Caulobacter crescentus	36,235	19-OCT-1999
		GB_PAT:I18647	3300	I18647	Sequence 6 from patent US 5500353.	Unknown.	36,821	07-OCT-1996
		GB_GSS13:AQ446197	751	AQ446197	nbxb0062D16r CUGI Rice BAC Library Oryza sativa genomic clone nbxb0062D16r, genomic survey sequence.	Oryza sativa	38,124	8-Apr-99
rxa00131	732	GB_BA1:MTY20B11	36330	Z95121	Mycobacterium tuberculosis H37Rv complete genome; segment 139/162.	Mycobacterium tuberculosis	43,571	17-Jun-98
		GB_BA1:SAR7932	15176	AJ007932	Streptomyces argillaceus mithramycin biosynthetic genes.	Streptomyces argillaceus	41,116	15-Jun-99
		GB_BA1:MTY20B11	36330	Z95121	Mycobacterium tuberculosis H37Rv complete genome; segment 139/162.	Mycobacterium tuberculosis	39,726	17-Jun-98
rxa00132	1557	GB_BA1:MTY20B11	36330	Z95121	Mycobacterium tuberculosis H37Rv complete genome; segment 139/162.	Mycobacterium tuberculosis	36,788	17-Jun-98
		GB_IN2:TVU40872	1882	U40872	Trichomonas vaginalis S-adenosyl-L-homocysteine hydrolase gene, complete cds.	Trichomonas vaginalis	61,914	31-OCT-1996
		GB_HTG6:AC010706	169265	AC010706	Drosophila melanogaster chromosome X clone BACR36D15 (D887) RPCI-98 36.D.15 map 13C-13E strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 74 unordered pieces.	Drosophila melanogaster	51,325	22-Nov-99
rxa00145	1059	GB_BA1:MTCY2B12	20431	Z81011	Mycobacterium tuberculosis H37Rv complete genome; segment 61/162.	Mycobacterium tuberculosis	63,365	18-Jun-98
		GB_BA1:PSEPYRBX	2273	L19649	Pseudomonas aeruginosa aspartate transcarbamoylase (pyrB) and dihydroorotase-like (pyrX) genes, complete cds's.	Pseudomonas aeruginosa	56,080	26-Jul-93
		GB_BA1:LLPYRBDNA	1468	X84262	L.leichmannii pyrB gene.	Lactobacillus leichmannii	47,514	29-Apr-97
rxa00146	1464	GB_BA1:MTCY2B12	20431	Z81011	Mycobacterium tuberculosis H37Rv complete genome; segment 61/162.	Mycobacterium tuberculosis	60,714	18-Jun-98
		GB_BA1:MTCY154	13935	Z98209	Mycobacterium tuberculosis H37Rv complete genome; segment 121/162.	Mycobacterium tuberculosis	39,229	17-Jun-98
		GB_BA1:MSGY154	40221	AD000002	Mycobacterium tuberculosis sequence from clone y154.	Mycobacterium tuberculosis	36,618	03-DEC-1996
rxa00147	1302	GB_BA1:MTCY2B12	20431	Z81011	Mycobacterium tuberculosis H37Rv complete genome; segment 61/162.	Mycobacterium tuberculosis	61,527	18-Jun-98
		GB_BA1:MSGB937CS	38914	L78820	Mycobacterium leprae cosmid B937 DNA sequence.	Mycobacterium leprae	59,538	15-Jun-96
		GB_BA1:PAU81259	7285	U81259	Pseudomonas aeruginosa dihydrodipicolinate reductase (dapB) gene, partial cds, carbamoylphosphate synthetase small subunit (carA) and carbamoylphosphate synthetase large subunit (carB) genes, complete cds, and FtsJ homolog (ftsJ) gene, partial cds.	Pseudomonas aeruginosa	55,396	23-DEC-1996
rxa00156	1233	GB_BA1:SC9B10	33320	AL009204	Streptomyces coelicolor cosmid 9B10.	Streptomyces coelicolor	52,666	10-Feb-99
		GB_BA2:AF002133	15437	AF002133	Mycobacterium avium strain GIR10 transcriptional regulator (mav81) gene, partial cds, aconitase (acn), invasin 1 (inv1), invasin 2 (inv2), transcriptional regulator (moxR), ketoacyl-reductase (fabG), enoyl-reductase (inhA) and ferrochelatase (mav272) genes, complete cds.	Mycobacterium avium	54,191	26-MAR-1998
		GB_BA1:D85417	7984	D85417	Propionibacterium freudenreichii hemY, hemH, hemB, hemX, hemR and hemL genes, complete cds.	Propionibacterium freudenreichii	46,667	6-Feb-99
rxa00166	783	GB_HTG3:AC008167	174223	AC008167	Homo sapiens clone NH0172O13, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	Homo sapiens	37,451	21-Aug-99
		GB_HTG3:AC008167	174223	AC008167	Homo sapiens clone NH0172O13, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	Homo sapiens	37,451	21-Aug-99
		GB_HTG4:AC010118	80605	AC010118	Drosophila melanogaster chromosome 3L/62B1 clone RPCI98-10D15, *** SEQUENCING IN PROGRESS ***, 51 unordered pieces.	Drosophila melanogaster	38,627	16-OCT-1999

TABLE 4: ALIGNMENT RESULTS

rxa00198	672	GB_BA1:AB024708	8734	AB024708	Corynebacterium glutamicum gltB and gltD genes for glutamine 2-oxoglutarate aminotransferase large and small subunits, complete cds.	Corynebacterium glutamicum	92,113	13-MAR-1999
		GB_BA1:AB024708	8734	AB024708	Corynebacterium glutamicum gltB and gltD genes for glutamine 2-oxoglutarate aminotransferase large and small subunits, complete cds.	Corynebacterium glutamicum	93,702	13-MAR-1999
		GB_EST24:AI232702	528	AI232702	EST229390 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone RKICF35 3' end, mRNA sequence.	Rattus sp.	34,221	31-Jan-99
rxa00216	1113	GB_HTG2:HSDJ850E9	117353	AL121758	Homo sapiens chromosome 20 clone RP5-850E9, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	37,965	03-DEC-1999
		GB_HTG2:HSDJ850E9	117353	AL121758	Homo sapiens chromosome 20 clone RP5-850E9, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	37,965	03-DEC-1999
		GB_PR2:CNS01DSA	159400	AL121766	Human chromosome 14 DNA sequence *** IN PROGRESS *** BAC R-412H8 of RPCI-11 library from chromosome 14 of Homo sapiens (Human), complete sequence.	Homo sapiens	38,796	11-Nov-99
rxa00219	1065	GB_HTG2:AC005079_0	110000	AC005079	Homo sapiens clone RG252P22, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	38,227	22-Nov-98
		GB_HTG2:AC005079_1	110000	AC005079	Homo sapiens clone RG252P22, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	38,227	22-Nov-98
		GB_HTG2:AC005079_1	110000	AC005079	Homo sapiens clone RG252P22, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	38,227	22-Nov-98
rxa00223	1212	GB_BA1:PPEA3NIF	19771	X99694	Plasmid pEA3 nitrogen fixation genes.	Enterobacter agglomerans	48,826	2-Aug-96
		GB_BA2:AF128444	2477	AF128444	Rhodobacter capsulatus molybdenum cofactor biosynthetic gene cluster, partial sequence.	Rhodobacter capsulatus	40,135	22-MAR-1999
		GB_HTG4:AC010111	138938	AC010111	Drosophila melanogaster chromosome 3L/70C1 clone RPCI98-9B18, *** SEQUENCING IN PROGRESS ***, 64 unordered pieces.	Drosophila melanogaster	39,527	16-OCT-1999
rxa00229	803	GB_BA2:AF124518	1758	AF124518	Corynebacterium glutamicum 3-dehydroquinase (aroD) and shikimate dehydrogenase (aroE) genes, complete cds.	Corynebacterium glutamicum	98,237	18-MAY-1999
		GB_PR3:AC004593	150221	AC004593	Homo sapiens PAC clone DJ0964C11 from 7p14-p15, complete sequence.	Homo sapiens	36,616	18-Apr-98
		GB_HTG2:AC006907	188972	AC006907	Caenorhabditis elegans clone Y76B12, *** SEQUENCING IN PROGRESS ***, 25 unordered pieces.	Caenorhabditis elegans	37,095	26-Feb-99
rxa00241	1626	GB_BA1:CGLYSI	4232	X60312	C.glutamicum lysI gene for L-lysine permease.	Corynebacterium glutamicum	100,000	30-Jan-92
		GB_HTG1:PFMAL13P1	192581	AL049180	Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Plasmodium falciparum	34,947	11-Aug-99
		GB_HTG1:PFMAL13P1	192581	AL049180	Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Plasmodium falciparum	34,947	11-Aug-99
rxa00262	1197	GB_IN2:EHU89655	3219	U89655	Entamoeba histolytica unconventional myosin IB mRNA, complete cds.	Entamoeba histolytica	36,496	23-MAY-1997
		GB_IN2:EHU89655	3219	U89655	Entamoeba histolytica unconventional myosin IB mRNA, complete cds.	Entamoeba histolytica	37,544	23-MAY-1997
rxa00266	531	GB_RO:AF016190	2939	AF016190	Mus musculus connexin-36 (Cx36) gene, complete cds.	Mus musculus	41,856	9-Feb-99
		EM_PAT:E09719	3505	E09719	DNA encoding precursor protein of alkaline cellulase.	Bacillus sp.	34,741	08-OCT-1997 (Rel. 52, Created)
		GB_PAT:E02133	3494	E02133	gDNA encoding alkaline cellulase.	Bacillus sp.	34,741	29-Sep-97
rxa00278	1155	GB_IN1:CELK05F6	36912	AF040653	Caenorhabditis elegans cosmid K05F6.	Caenorhabditis elegans	36,943	6-Jan-98

TABLE 4: ALIGNMENT RESULTS

		GB_BA1:CGU43535	2531	U43535	Corynebacterium glutamicum multidrug resistance protein (cmr) gene, complete cds.	Corynebacterium glutamicum	36,658	9-Apr-97
		GB_RO:RNU30789	3510	U30789	Rattus norvegicus clone N27 mRNA.	Rattus norvegicus	38,190	20-Aug-96
rxa00295	1125	GB_BA2:CGU31281	1614	U31281	Corynebacterium glutamicum biotin synthase (bioB) gene, complete cds.	Corynebacterium glutamicum	99,111	21-Nov-96
		GB_BA1:BRLBIOBA	1647	D14084	Brevibacterium flavum gene for biotin synthetase, complete cds.	Corynebacterium glutamicum	98,489	3-Feb-99
		GB_PAT:E03937	1005	E03937	DNA sequence encoding Brevibacterium flavum biotin-synthase.	Corynebacterium glutamicum	98,207	29-Sep-97
rxa00323	1461	GB_BA1:MTCY427	38110	Z70692	Mycobacterium tuberculosis H37Rv complete genome; segment 99/162.	Mycobacterium tuberculosis	35,615	24-Jun-99
		GB_BA1:MSGB32CS	36404	L78818	Mycobacterium leprae cosmid B32 DNA sequence.	Mycobacterium leprae	60,917	15-Jun-96
		GB_BA1:MTCY427	38110	Z70692	Mycobacterium tuberculosis H37Rv complete genome; segment 99/162.	Mycobacterium tuberculosis	44,606	24-Jun-99
rxa00324	3258	GB_BA1:MSGB32CS	36404	L78818	Mycobacterium leprae cosmid B32 DNA sequence.	Mycobacterium leprae	52,516	15-Jun-96
		GB_BA1:MTCY427	38110	Z70692	Mycobacterium tuberculosis H37Rv complete genome; segment 99/162.	Mycobacterium tuberculosis	38,079	24-Jun-99
		GB_OM:BOVELA	3242	J02717	Bovine elastin a mRNA, complete cds.	Bos taurus	39,351	27-Apr-93
rxa00330	1566	GB_BA1:CGTHRC	3120	X56037	Corynebacterium glutamicum thrC gene for threonine synthase (EC 4.2.99.2).	Corynebacterium glutamicum	99,808	17-Jun-97
		GB_PAT:I09078	3146	I09078	Sequence 4 from Patent WO 8809819.	Unknown.	99,617	02-DEC-1994
		GB_BA1:BLTHRESYN	1892	Z29563	Brevibacterium lactofermentum; ATCC 13869;; DNA (genomic);.	Corynebacterium glutamicum	99,170	20-Sep-95
rxa00335	1554	GB_BA1:CGGLNA	3686	Y13221	Corynebacterium glutamicum glnA gene.	Corynebacterium glutamicum	100,000	28-Aug-97
		GB_BA2:AF005635	1690	AF005635	Corynebacterium glutamicum glutamine synthetase (glnA) gene, complete cds.	Corynebacterium glutamicum	98,906	14-Jun-99
		GB_BA1:MSGB27CS	38793	L78817	Mycobacterium leprae cosmid B27 DNA sequence.	Mycobacterium leprae	66,345	15-Jun-96
rxa00347	891	GB_EST27:AI455217	624	AI455217	LD21828.3prime LD Drosophila melanogaster embryo pOT2 Drosophila melanogaster cDNA clone LD21828 3prime, mRNA sequence.	Drosophila melanogaster	34,510	09-MAR-1999
		GB_BA2:SSU30252	2891	U30252	Synechococcus PCC7942 nucleoside diphosphate kinase and ORF2 protein genes, complete cds, ORF1 protein gene, partial cds, and neutral site I for vector use.	Synechococcus PCC7942	37,084	29-OCT-1999
		GB_EST21:AA911262	581	AA911262	oe75a02.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1417418 3' similar to gb:A18757 UROKINASE PLASMINOGEN ACTIVATOR SURFACE RECEPTOR, GPI-ANCHORED (HUMAN);, mRNA sequence.	Homo sapiens	37,500	21-Apr-98
rxa00351	1578	GB_BA1:MLU15187	36138	U15187	Mycobacterium leprae cosmid L296.	Mycobacterium leprae	52,972	09-MAR-1995
		GB_IN2:AC004373	72722	AC004373	Drosophila melanogaster DNA sequence (P1 DS05273 (D80)), complete sequence.	Drosophila melanogaster	46,341	17-Jul-98
		GB_IN2:AF145653	3197	AF145653	Drosophila melanogaster clone GH08860 BcDNA.GH08860 (BcDNA.GH08860) mRNA, complete cds.	Drosophila melanogaster	49,471	14-Jun-99
rxa00365	727	GB_BA1:AB024708	8734	AB024708	Corynebacterium glutamicum gltB and gltD genes for glutamine 2-oxoglutarate aminotransferase large and small subunits, complete cds.	Corynebacterium glutamicum	96,556	13-MAR-1999
		GB_BA1:MTCY1A6	37751	Z83864	Mycobacterium tuberculosis H37Rv complete genome; segment 159/162.	Mycobacterium tuberculosis	39,496	17-Jun-98
		GB_BA1:SC3A3	15901	AL109849	Streptomyces coelicolor cosmid 3A3.	Streptomyces coelicolor A3(2)	37,946	16-Aug-99
rxa00366	480	GB_BA1:AB024708	8734	AB024708	Corynebacterium glutamicum gltB and gltD genes for glutamine 2-oxoglutarate aminotransferase large and small subunits, complete cds.	Corynebacterium glutamicum	99,374	13-MAR-1999
		GB_BA1:MTCY1A6	37751	Z83864	Mycobacterium tuberculosis H37Rv complete genome; segment 159/162.	Mycobacterium tuberculosis	41,333	17-Jun-98
		GB_BA1:SC3A3	15901	AL109849	Streptomyces coelicolor cosmid 3A3.	Streptomyces coelicolor A3(2)	37,554	16-Aug-99

TABLE 4: ALIGNMENT RESULTS

rxa00367	4653	GB_BA1:AB024708	8734	AB024708	Corynebacterium glutamicum gltB and gltD genes for glutamine 2-oxoglutarate aminotransferase large and small subunits, complete cds.	Corynebacterium glutamicum	99,312	13-MAR-1999
		GB_BA1:MTCY1A6	37751	Z83864	Mycobacterium tuberculosis H37Rv complete genome; segment 159/162.	Mycobacterium tuberculosis	36,971	17-Jun-98
		GB_BA1:SC3A3	15901	AL109849	Streptomyces coelicolor cosmid 3A3.	Streptomyces coelicolor A3(2)	37,905	16-Aug-99
rxa00371	1917	GB_VI:SBVORFS	7568	M89923	Sugarcane bacilliform virus ORF 1,2,and 3 DNA, complete cds.	Sugarcane bacilliform virus	35,843	12-Jun-93
		GB_EST37:AI967505	380	AI967505	Ljirnpest03-215-c10 Ljirnp Lambda HybriZap two-hybrid library Lotus japonicus cDNA clone LP215-03-c10 5' similar to 60S ribosomal protein L39, mRNA sequence.	Lotus japonicus	42,593	24-Aug-99
		GB_IN1:CELK09H9	37881	AF043700	Caenorhabditis elegans cosmid K09H9.	Caenorhabditis elegans	34,295	22-Jan-98
rxa00377	1245	GB_BA1:CCU13664	1678	U13664	Caulobacter crescentus uroporphyrinogen decarboxylase homolog (hemE) gene, partial cds.	Caulobacter crescentus	36,832	24-MAR-1995
		GB_PL1:ANSDGENE	1299	Y08866	A.nidulans sD gene.	Emericella nidulans	39,603	17-OCT-1996
		GB_GSS4:AQ730303	483	AQ730303	HS_5505_B1_C04_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1081 Col=7 Row=F, genomic survey sequence.	Homo sapiens	36,728	15-Jul-99
rxa00382	1425	GB_BA1:PAHEML	4444	X82072	P.aeruginosa hemL gene.	Pseudomonas aeruginosa	54,175	18-DEC-1995
		GB_BA1:MTY25D10	40838	Z95558	Mycobacterium tuberculosis H37Rv complete genome; segment 28/162.	Mycobacterium tuberculosis	61,143	17-Jun-98
		GB_BA1:MSGY224	40051	AD000004	Mycobacterium tuberculosis sequence from clone y224.	Mycobacterium tuberculosis	61,143	03-DEC-1996
rxa00383	1467	GB_BA1:MLCB1222	34714	AL049491	Mycobacterium leprae cosmid B1222.	Mycobacterium leprae	43,981	27-Aug-99
		GB_HTG2:AC006269	167171	AC006269	Homo sapiens chromosome 17 clone hRPK.515_E_23 map 17, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.	Homo sapiens	35,444	10-Jun-99
		GB_HTG2:AC007638	178053	AC007638	Homo sapiens chromosome 17 clone hRPK.515_O_17 map 17, *** SEQUENCING IN PROGRESS ***, 8 unordered pieces.	Homo sapiens	34,821	22-MAY-1999
rxa00391	843	GB_EST38:AW017053	613	AW017053	EST272398 Schistosoma mansoni male, Phil LoVerde/Joe Merrick Schistosoma mansoni cDNA clone SMMAS14 5' end, mRNA sequence.	Schistosoma mansoni	40,472	10-Sep-99
		GB_PAT:AR065852	32207	AR065852	Sequence 20 from patent US 5849564.	Unknown.	38,586	29-Sep-99
		GB_VI:AF148805	28559	AF148805	Kaposi's sarcoma-associated herpesvirus ORF 68 gene, partial cds; and ORF 69, kaposin, v-FLIP, v-cyclin, latent nuclear antigen, ORF K14, v-GPCR, putative phosphoribosylformylglycinamidine synthase, and LAMP (LAMP) genes, complete cds.	Kaposi's sarcoma-associated herpesvirus	38,509	2-Aug-99
rxa00393	1017	GB_BA1:MTY25D10	40838	Z95558	Mycobacterium tuberculosis H37Rv complete genome; segment 28/162.	Mycobacterium tuberculosis	36,308	17-Jun-98
		GB_BA1:MSGY224	40051	AD000004	Mycobacterium tuberculosis sequence from clone y224.	Mycobacterium tuberculosis	39,282	03-DEC-1996
		GB_BA1:MLB1306	7762	Y13803	Mycobacterium leprae cosmid B1306 DNA.	Mycobacterium leprae	39,228	24-Jun-97
rxa00402	623	GB_BA2:AF052652	2096	AF052652	Corynebacterium glutamicum homoserine O-acetyltransferase (metA) gene, complete cds.	Corynebacterium glutamicum	99,672	19-MAR-1998
		GB_BA2:AF109162	4514	AF109162	Corynebacterium diphtheriae heme uptake locus, complete sequence.	Corynebacterium diphtheriae	40,830	8-Jun-99
		GB_BA2:AF092918	20758	AF092918	Pseudomonas alcaligenes outer membrane Xcp-secretion system gene cluster.	Pseudomonas alcaligenes	50,161	06-DEC-1998
rxa00403	1254	GB_BA2:AF052652	2096	AF052652	Corynebacterium glutamicum homoserine O-acetyltransferase (metA) gene, complete cds.	Corynebacterium glutamicum	99,920	19-MAR-1998
		GB_BA1:MTV016	53662	AL021841	Mycobacterium tuberculosis H37Rv complete genome; segment 143/162.	Mycobacterium tuberculosis	52,898	23-Jun-99
		GB_EST23:AI111288	750	AI111288	SWOvAMCAQ02A05SK Onchocerca volvulus adult male cDNA (SAW98MLW-OvAM) Onchocerca volvulus cDNA clone SWOvAMCAQ02A05 5', mRNA sequence.	Onchocerca volvulus	37,565	31-Aug-98

TABLE 4: ALIGNMENT RESULTS

rxa00405	613	GB_BA1:MTV016	53662	AL021841	Mycobacterium tuberculosis H37Rv complete genome; segment 143/162.	Mycobacterium tuberculosis	57,259	23-Jun-99
		GB_PR4:AC005145	143678	AC005145	Homo sapiens Xp22-166-169 GSHB-523A23 (Genome Systems Human BAC library) complete sequence.	Homo sapiens	34,179	08-DEC-1998
		GB_BA1:MTV016	53662	AL021841	Mycobacterium tuberculosis H37Rv complete genome; segment 143/162.	Mycobacterium tuberculosis	40,169	23-Jun-99
rxa00420	1587	GB_BA1:MTY13D12	37085	Z80343	Mycobacterium tuberculosis H37Rv complete genome; segment 156/162.	Mycobacterium tuberculosis	62,031	17-Jun-98
		GB_BA1:MSGY126	37164	AD000012	Mycobacterium tuberculosis sequence from clone y126.	Mycobacterium tuberculosis	61,902	10-DEC-1996
		GB_BA1:MSGB971CS	37566	L78821	Mycobacterium leprae cosmid B971 DNA sequence.	Mycobacterium leprae	39,651	15-Jun-96
rxa00435	1296	GB_BA1:AFACBBTZ	2760	M68904	Alcaligenes eutrophus chromsomal transketolase (cbbTc) and phosphoglycolate phosphatase (cbbZc) genes, complete cds.	Ralstonia eutropha	38,677	27-Jul-94
		GB_HTG4:AC009541	169583	AC009541	Homo sapiens chromosome 7, *** SEQUENCING IN PROGRESS ***, 25 unordered pieces.	Homo sapiens	36,335	12-OCT-1999
		GB_HTG4:AC009541	169583	AC009541	Homo sapiens chromosome 7, *** SEQUENCING IN PROGRESS ***, 25 unordered pieces.	Homo sapiens	36,335	12-OCT-1999
rxa00437	579	GB_PR4:AC005951	155450	AC005951	Homo sapiens chromosome 17, clone hRPK.372_K_20, complete sequence.	Homo sapiens	31,738	18-Nov-98
		GB_BA1:SC2A11	22789	AL031184	Streptomyces coelicolor cosmid 2A11.	Streptomyces coelicolor	43,262	5-Aug-98
		GB_PR4:AC005951	155450	AC005951	Homo sapiens chromosome 17, clone hRPK.372_K_20, complete sequence.	Homo sapiens	37,647	18-Nov-98
rxa00439	591	GB_BA1:MTV016	53662	AL021841	Mycobacterium tuberculosis H37Rv complete genome; segment 143/162.	Mycobacterium tuberculosis	37,088	23-Jun-99
		GB_PL2:AF167358	1022	AF167358	Rumex acetosa expansin (EXP3) gene, partial cds.	Rumex acetosa	46,538	17-Aug-99
		GB_HTG3:AC009120	269445	AC009120	Homo sapiens chromosome 16 clone RPCI-11_484E3, *** SEQUENCING IN PROGRESS ***, 34 unordered pieces.	Homo sapiens	43,276	3-Aug-99
rxa00440	582	GB_BA2:SKZ86111	7860	Z86111	Streptomyces lividans rpsP, trmD, rplS, sipW, sipX, sipY, sipZ, mutT genes and 4 open reading frames.	Streptomyces lividans	43,080	27-OCT-1999
		GB_BA1:SC2E1	38962	AL023797	Streptomyces coelicolor cosmid 2E1.	Streptomyces coelicolor	42,931	4-Jun-98
		GB_BA1:SC2E1	38962	AL023797	Streptomyces coelicolor cosmid 2E1.	Streptomyces coelicolor	36,702	4-Jun-98
rxa00441	1287	GB_PR2:HS173D1	117338	AL031984	Human DNA sequence from clone 173D1 on chromosome 1p36.21-36.33.Contains ESTs, STSs and GSSs, complete sequence.	Homo sapiens	38,027	23-Nov-99
		GB_HTG2:HSDJ719K3	267114	AL109931	Homo sapiens chromosome X clone RP4-719K3 map q21.1-21.31, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	34,521	03-DEC-1999
		GB_HTG2:HSDJ719K3	267114	AL109931	Homo sapiens chromosome X clone RP4-719K3 map q21.1-21.31, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	34,521	03-DEC-1999
rxa00446	987	GB_BA1:SCD78	36224	AL034355	Streptomyces coelicolor cosmid D78.	Streptomyces coelicolor	56,410	26-Nov-98
		GB_HTG4:AC009367	226055	AC009367	Drosophila melanogaster chromosome 3L/76A2 clone RPCI98-48B15, *** SEQUENCING IN PROGRESS ***, 44 unordered pieces.	Drosophila melanogaster	34,959	16-OCT-1999
		GB_HTG4:AC009367	226055	AC009367	Drosophila melanogaster chromosome 3L/76A2 clone RPCI98-48B15, *** SEQUENCING IN PROGRESS ***, 44 unordered pieces.	Drosophila melanogaster	34,959	16-OCT-1999
rxa00448	1143	GB_PR3:AC003670	88945	AC003670	Homo sapiens 12q13.1 PAC RPCI1-130F5 (Roswell Park Cancer Institute Human PAC library) complete sequence.	Homo sapiens	35,682	9-Jun-98
		GB_HTG2:AF029367	148676	AF029367	Homo sapiens chromosome 12 clone RPCI-1 130F5 map 12q13.1, *** SEQUENCING IN PROGRESS ***, 156 unordered pieces.	Homo sapiens	31,373	18-OCT-1997
		GB_HTG2:AF029367	148676	AF029367	Homo sapiens chromosome 12 clone RPCI-1 130F5 map 12q13.1, *** SEQUENCING IN PROGRESS ***, 156 unordered pieces.	Homo sapiens	31,373	18-OCT-1997

TABLE 4: ALIGNMENT RESULTS

rxa00450	424	GB_HTG2:AC007824	133361	AC007824	Drosophila melanogaster chromosome 3 clone BACR02L16 (D715) RPCI-98 02.L.16 map 89E-90A strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 91 unordered pieces.	Drosophila melanogaster	40,000	2-Aug-99
		GB_HTG2:AC007824	133361	AC007824	Drosophila melanogaster chromosome 3 clone BACR02L16 (D715) RPCI-98 02.L.16 map 89E-90A strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 91 unordered pieces.	Drosophila melanogaster	40,000	2-Aug-99
		GB_EST35:AI818057	412	AI818057	wk14a08.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2412278 3' similar to gb:Y00764 UBIQUINOL-CYTOCHROME C REDUCTASE 11 KD PROTEIN (HUMAN);, mRNA sequence.	Homo sapiens	35,714	24-Aug-99
rxa00461	975	GB_BA1:MLCB1779	43254	Z98271	Mycobacterium leprae cosmid B1779.	Mycobacterium leprae	39,308	8-Aug-97
		GB_IN1:DMC86E4	29352	AL021086	Drosophila melanogaster cosmid clone 86E4.	Drosophila melanogaster	37,487	27-Apr-99
		GB_GSS15:AQ640325	467	AQ640325	927P1-2H3.TP 927P1 Trypanosoma brucei genomic clone 927P1-2H3, genomic survey sequence.	Trypanosoma brucei	38,116	8-Jul-99
rxa00465								
rxa00487	1692	GB_BA1:BAGUAA	3866	Y10499	B.ammoniagenes guaA gene.	Corynebacterium ammoniagenes	74,259	8-Jan-98
		GB_BA2:U00015	42325	U00015	Mycobacterium leprae cosmid B1620.	Mycobacterium leprae	37,248	01-MAR-1994
		GB_BA1:MTCY78	33818	Z77165	Mycobacterium tuberculosis H37Rv complete genome; segment 145/162.	Mycobacterium tuberculosis	39,725	17-Jun-98
rxa00488	1641	GB_BA1:MTCY78	33818	Z77165	Mycobacterium tuberculosis H37Rv complete genome; segment 145/162.	Mycobacterium tuberculosis	39,451	17-Jun-98
		GB_BA2:U00015	42325	U00015	Mycobacterium leprae cosmid B1620.	Mycobacterium leprae	39,178	01-MAR-1994
		GB_BA1:SCAJ10601	4692	AJ010601	Streptomyces coelicolor A3(2) DNA for whiD and whiK loci.	Streptomyces coelicolor	60,835	17-Sep-98
rxa00489	1245	GB_BA2:U00015	42325	U00015	Mycobacterium leprae cosmid B1620.	Mycobacterium leprae	38,041	01-MAR-1994
		GB_HTG2:HS225E12	126464	AL031772	Homo sapiens chromosome 6 clone RP1-225E12 map q24, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	36,756	03-DEC-1999
		GB_HTG2:HS225E12	126464	AL031772	Homo sapiens chromosome 6 clone RP1-225E12 map q24, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	36,756	03-DEC-1999
rxa00533	1155	GB_BA1:CGLYS	2803	X57226	C. glutamicum lysC-alpha, lysC-beta and asd genes for aspartokinase-alpha and -beta subunits, and aspartate beta semialdehyde dehydrogenase, respectively (EC 2.7.2.4; EC 1.2.1.11).	Corynebacterium glutamicum	99,913	17-Feb-97
		GB_BA1:CGCYSCASD	1591	X82928	C.glutamicum aspartate-semialdehyde dehydrogenase gene.	Corynebacterium glutamicum	99,221	17-Feb-97
		GB_PAT:A07546	2112	A07546	Recombinant DNA fragment (PstI-XhoI).	synthetic construct	99,391	30-Jul-93
rxa00534	1386	GB_BA1:CGLYS	2803	X57226	C. glutamicum lysC-alpha, lysC-beta and asd genes for aspartokinase-alpha and -beta subunits, and aspartate beta semialdehyde dehydrogenase, respectively (EC 2.7.2.4; EC 1.2.1.11).	Corynebacterium glutamicum	99,856	17-Feb-97
		GB_BA1:CORASKD	2957	L16848	Corynebacterium flavum aspartokinase (ask), and aspartate-semialdehyde dehydrogenase (asd) genes, complete cds.	Corynebacterium flavescens	98,701	11-Jun-93
		GB_PAT:E14514	1643	E14514	DNA encoding Brevibacterium aspartokinase.	Corynebacterium glutamicum	98,773	28-Jul-99
rxa00536	1494	GB_BA1:CGLEUA	3492	X70959	C.glutamicum gene leuA for isopropylmalate synthase.	Corynebacterium glutamicum	100,000	10-Feb-99
		GB_BA1:MTV025	121125	AL022121	Mycobacterium tuberculosis H37Rv complete genome; segment 155/162.	Mycobacterium tuberculosis	68,003	24-Jun-99
		GB_BA1:MTU88526	2412	U88526	Mycobacterium tuberculosis putative alpha-isopropyl malate synthase (leuA) gene, complete cds.	Mycobacterium tuberculosis	68,185	26-Feb-97

TABLE 4: ALIGNMENT RESULTS

rxa00537	2409	GB_BA2:SCD25	41622	AL118514	Streptomyces coelicolor cosmid D25.	Streptomyces coelicolor A3(2)	63,187	21-Sep-99
		GB_BA1:MTCY7H7A	10451	Z95618	Mycobacterium tuberculosis H37Rv complete genome; segment 39/162.	Mycobacterium tuberculosis	62,401	17-Jun-98
		GB_BA1:MTU34956	2462	U34956	Mycobacterium tuberculosis phosphoribosylformylglycinamidine synthase (purL) gene, complete cds.	Mycobacterium tuberculosis	62,205	28-Jan-97
rxa00541	792	GB_PAT:I92052	2115	I92052	Sequence 19 from patent US 5726299.	Unknown.	98,359	01-DEC-1998
		GB_BA1:MLCB5	38109	Z95151	Mycobacterium leprae cosmid B5.	Mycobacterium leprae	62,468	24-Jun-97
		GB_BA1:MTCY369	36850	Z80226	Mycobacterium tuberculosis H37Rv complete genome; segment 36/162.	Mycobacterium tuberculosis	60,814	17-Jun-98
rxa00558	1470	GB_BA1:BAPURF	1885	X91252	B.ammoniagenes purF gene.	Corynebacterium ammoniagenes	66,095	5-Jun-97
		GB_BA1:MLU15182	40123	U15182	Mycobacterium leprae cosmid B2266.	Mycobacterium leprae	64,315	09-MAR-1995
		GB_BA1:MTCY7H7A	10451	Z95618	Mycobacterium tuberculosis H37Rv complete genome; segment 39/162.	Mycobacterium tuberculosis	64,863	17-Jun-98
rxa00579	1983	GB_PAT:AR016483	2104	AR016483	Sequence 1 from patent US 5776740.	Unknown.	98,810	05-DEC-1998
		EM_PAT:E11273	2104	E11273	DNA encoding serine hydroxymethyl transferase.	Corynebacterium glutamicum	98,810	08-OCT-1997 (Rel. 52, Created)
		GB_PAT:E12594	2104	E12594	DNA encoding serine hydroxymethyltransferase from Brevibacterium flavum.	Corynebacterium glutamicum	98,810	24-Jun-98
rxa00580	1425	GB_PAT:E12594	2104	E12594	DNA encoding serine hydroxymethyltransferase from Brevibacterium flavum.	Corynebacterium glutamicum	99,368	24-Jun-98
		GB_PAT:AR016483	2104	AR016483	Sequence 1 from patent US 5776740.	Unknown.	99,368	05-DEC-1998
		EM_PAT:E11273	2104	E11273	DNA encoding serine hydroxymethyl transferase.	Corynebacterium glutamicum	99,368	08-OCT-1997 (Rel. 52, Created)
rxa00581	1092	GB_PAT:E12594	2104	E12594	DNA encoding serine hydroxymethyltransferase from Brevibacterium flavum.	Corynebacterium glutamicum	37,071	24-Jun-98
		EM_PAT:E11273	2104	E11273	DNA encoding serine hydroxymethyl transferase.	Corynebacterium glutamicum	37,071	08-OCT-1997 (Rel. 52, Created)
		GB_PAT:AR016483	2104	AR016483	Sequence 1 from patent US 5776740.	Unknown.	37,071	05-DEC-1998
rxa00584	1248	GB_BA1:CORAHPS	2570	L07603	Corynebacterium glutamicum 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene, complete cds.	Corynebacterium glutamicum	98,236	26-Apr-93
		GB_BA1:AOPCZA361	37941	AJ223998	Amycolatopsis orientalis cosmid PCZA361.	Amycolatopsis orientalis	54,553	29-MAR-1999
		GB_BA1:D90714	14358	D90714	Escherichia coli genomic DNA. (16.8 - 17.1 min).	Escherichia coli	53,312	7-Feb-99
rxa00618	1230	GB_EST19:AA802737	280	AA802737	GM06236.5prime GM Drosophila melanogaster ovary BlueScript Drosophila melanogaster cDNA clone GM06236 5prime, mRNA sequence.	Drosophila melanogaster	39,928	25-Nov-98
		GB_EST28:AI534381	581	AI534381	SD07186.5prime SD Drosophila melanogaster Schneider L2 cell culture pOT2 Drosophila melanogaster cDNA clone SD07186 5prime similar to X89858: Ani FBgn0011558 PID:g927407 SPTREMBL:Q24240, mRNA sequence.	Drosophila melanogaster	41,136	18-MAR-1999
		GB_IN1:DMANILLIN	4029	X89858	D.melanogaster mRNA for anillin protein.	Drosophila melanogaster	34,398	8-Nov-95
rxa00619	1551	GB_BA1:MTCY369	36850	Z80226	Mycobacterium tuberculosis H37Rv complete genome; segment 36/162.	Mycobacterium tuberculosis	62,776	17-Jun-98
		GB_BA1:MLCB5	38109	Z95151	Mycobacterium leprae cosmid B5.	Mycobacterium leprae	61,831	24-Jun-97
		GB_PAT:A60305	1845	A60305	Sequence 5 from Patent WO9708323.	unidentified	61,785	06-MAR-1998
rxa00620	1014	GB_PL2:AF063247	1450	AF063247	Pneumocystis carinii f. sp. ratti enolase mRNA, complete cds.	Pneumocystis carinii f. sp. ratti	41,060	5-Jan-99
		GB_BA1:STMAPP	2069	M91546	Streptomyces lividans aminopeptidase P (PepP) gene, complete cds.	Streptomyces lividans	37,126	12-Jun-93

TABLE 4: ALIGNMENT RESULTS

		GB_HTG3:AC008763	214575	AC008763	Homo sapiens chromosome 19 clone CITB-E1_3214H19, *** SEQUENCING IN PROGRESS ***, 21 unordered pieces.	Homo sapiens	40,020	3-Aug-99
rxa00624	810	GB_IN1:CEY41E3	150641	Z95559	Caenorhabditis elegans cosmid Y41E3, complete sequence.	Caenorhabditis elegans	36,986	2-Sep-99
		GB_EST13:AA362167	372	AA362167	EST71561 Macrophage I Homo sapiens cDNA 5' end, mRNA sequence.	Homo sapiens	38,378	21-Apr-97
		GB_IN1:CEY41E3	150641	Z95559	Caenorhabditis elegans cosmid Y41E3, complete sequence.	Caenorhabditis elegans	37,694	2-Sep-99
rxa00626	1386	GB_BA1:MTCY369	36850	Z80226	Mycobacterium tuberculosis H37Rv complete genome; segment 36/162.	Mycobacterium tuberculosis	57,971	17-Jun-98
		GB_BA1:MLCB5	38109	Z95151	Mycobacterium leprae cosmid B5.	Mycobacterium leprae	58,806	24-Jun-97
		GB_BA1:MLU15187	36138	U15187	Mycobacterium leprae cosmid L296.	Mycobacterium leprae	38,007	09-MAR-1995
rxa00632	795	GB_BA1:BRLBIOAD	2272	D14083	Brevibacterium flavum genes for 7,8-diaminopelargonic acid aminotransferase and dethiobiotin synthetase, complete cds.	Corynebacterium glutamicum	97,358	3-Feb-99
		GB_PAT:E04041	675	E04041	DNA sequence coding for desthiobiotinsynthetase.	Corynebacterium glutamicum	98,074	29-Sep-97
		GB_PAT:E04040	1272	E04040	DNA sequence coding for diamino pelargonic acid aminotransferase.	Corynebacterium glutamicum	93,814	29-Sep-97
rxa00633	1392	GB_BA1:BRLBIOAD	2272	D14083	Brevibacterium flavum genes for 7,8-diaminopelargonic acid aminotransferase and dethiobiotin synthetase, complete cds.	Corynebacterium glutamicum	95,690	3-Feb-99
		GB_PAT:E04040	1272	E04040	DNA sequence coding for diamino pelargonic acid aminotransferase.	Corynebacterium glutamicum	95,755	29-Sep-97
		GB_BA2:EHU38519	1290	U38519	Erwinia herbicola adenosylmethionine-8-amino-7-oxononanoate transaminase (bioA) gene, complete cds.	Erwinia herbicola	55,564	4-Nov-96
rxa00688	666	GB_BA1:MTV041	28826	AL021958	Mycobacterium tuberculosis H37Rv complete genome; segment 35/162.	Mycobacterium tuberculosis	60,030	17-Jun-98
		GB_BA1:BRLSECY	1516	D14162	Brevibacterium flavum gene for SecY protein (complete cds) and gene or adenylate kinase (partial cds).	Corynebacterium glutamicum	99,563	3-Feb-99
		GB_BA2:MBU77912	7163	U77912	Mycobacterium bovis MBE50a gene, partial cds; and MBE50b, MBE50c, preprotein translocase SecY subunit (secY), adenylate kinase (adk), methionine aminopeptidase (map), RNA polymerase ECF sigma factor (sigE50), MBE50d, and MBE50e genes, complete cds.	Mycobacterium bovis	60,030	27-Jan-99
rxa00708	930	GB_BA2:AF157493	25454	AF157493	Zymomonas mobilis ZM4 fosmid clone 42D7, complete sequence.	Zymomonas mobilis	39,116	5-Jul-99
		GB_PAT:I00836	1853	I00836	Sequence 1 from Patent US 4758514.	Unknown.	47,419	21-MAY-1993
		GB_PAT:E00311	1853	E00311	DNA coding of 2,5-diketogluconic acid reductase.	unidentified	47,419	29-Sep-97
rxa00717	1083	GB_PAT:I78753	1187	I78753	Sequence 9 from patent US 5693781.	Unknown.	37,814	3-Apr-98
		GB_PAT:I92042	1187	I92042	Sequence 9 from patent US 5726299.	Unknown.	37,814	01-DEC-1998
		GB_BA1:MTCI125	37432	Z98268	Mycobacterium tuberculosis H37Rv complete genome; segment 76/162.	Mycobacterium tuberculosis	50,647	17-Jun-98
rxa00718	831	GB_BA1:MTCI125	37432	Z98268	Mycobacterium tuberculosis H37Rv complete genome; segment 76/162.	Mycobacterium tuberculosis	55,228	17-Jun-98
		GB_BA1:MTCI125	37432	Z98268	Mycobacterium tuberculosis H37Rv complete genome; segment 76/162.	Mycobacterium tuberculosis	40,300	17-Jun-98
		GB_GSS12:AQ420755	671	AQ420755	RPCI-11-168G18.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-168G18, genomic survey sequence.	Homo sapiens	35,750	23-MAR-1999
rxa00727	1035	GB_HTG3:AC008332	118545	AC008332	Drosophila melanogaster chromosome 2 clone BACR48D10 (D867) RPCI-98 48.D.10 map 34A-34A strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 78 unordered pieces.	Drosophila melanogaster	40,634	6-Aug-99
		GB_HTG3:AC008332	118545	AC008332	Drosophila melanogaster chromosome 2 clone BACR48D10 (D867) RPCI-98 48.D.10 map 34A-34A strain y; cn bw sp, *** SEQUENCING IN PROGRESS***, 78 unordered pieces.	Drosophila melanogaster	40,634	6-Aug-99
		GB_HTG3:AC008332	118545	AC008332	Drosophila melanogaster chromosome 2 clone BACR48D10 (D867) RPCI-98 48.D.10 map 34A-34A strain y; cn bw sp, *** SEQUENCING IN PROGRESS***, 78 unordered pieces.	Drosophila melanogaster	33,888	6-Aug-99

TABLE 4: ALIGNMENT RESULTS

rxa00766	966	GB_HTG2:AC006789	83823	AC006789	Caenorhabditis elegans clone Y49F6, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.	Caenorhabditis elegans	36,737	25-Feb-99
		GB_HTG2:AC006789	83823	AC006789	Caenorhabditis elegans clone Y49F6, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.	Caenorhabditis elegans	36,737	25-Feb-99
		GB_BA1:D90810	20476	D90810	E.coli genomic DNA, Kohara clone #319(37.4-37.8 min.).	Escherichia coli	36,526	29-MAY-1997
rxa00770	1293	GB_BA1:MTV043	68848	AL022004	Mycobacterium tuberculosis H37Rv complete genome; segment 40/162.	Mycobacterium tuberculosis	66,193	24-Jun-99
		GB_BA1:MLU15182	40123	U15182	Mycobacterium leprae cosmid B2266.	Mycobacterium leprae	61,443	09-MAR-1995
		GB_BA2:SCD25	41622	AL118514	Streptomyces coelicolor cosmid D25.	Streptomyces coelicolor A3(2)	59,938	21-Sep-99
rxa00779	1056	GB_HTG1:CER08A5	51920	Z82281	Caenorhabditis elegans chromosome V clone R08A5, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	64,896	14-OCT-1998
		GB_HTG1:CER08A5	51920	Z82281	Caenorhabditis elegans chromosome V clone R08A5, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	64,896	14-OCT-1998
		GB_PL2:AF078693	1492	AF078693	Chlamydomonas reinhardtii putative O-acetylserine(thiol)lyase precursor (Crcys-1A) mRNA, nuclear gene encoding organellar protein, complete cds.	Chlamydomonas reinhardtii	57,970	3-Nov-99
rxa00780	669	GB_BA1:MTCY98	31225	Z83860	Mycobacterium tuberculosis H37Rv complete genome; segment 103/162.	Mycobacterium tuberculosis	54,410	17-Jun-98
		GB_BA1:AVINIFREG	7099	M60090	Azotobacter chroococcum nifU, nifS, nifV, nifP, nifW, nifZ and nifM genes, complete cds.	Azotobacter chroococcum	51,729	26-Apr-93
		GB_BA2:AF001780	6701	AF001780	Cyanothece PCC 8801 NifP (nifP), nitrogenase (nifB), FdxN (fdxN), NifS (nifS) and NifU (nifU) genes, complete cds, and NifH (nifH) gene, partial cds.	Cyanothece PCC8801	36,309	08-MAR-1999
rxa00838	1023	GB_EST1:Z30506	329	Z30506	ATTS2430 AC16H Arabidopsis thaliana cDNA clone TAI306 3', mRNA sequence.	Arabidopsis thaliana	44,308	11-MAR-1994
		GB_PL2:AC006258	110469	AC006258	Arabidopsis thaliana BAC F18G18 from chromosome V near 60.5 cM, complete sequence.	Arabidopsis thaliana	35,571	28-DEC-1998
		GB_EST37:AI998439	455	AI998439	701545695 A. thaliana, Columbia Col-0, rosette-2 Arabidopsis thaliana cDNA clone 701545695, mRNA sequence.	Arabidopsis thaliana	36,044	8-Sep-99
rxa00863	867	GB_BA1:BLDAPAB	3572	Z21502	B.lactofermentum dapA and dapB genes for dihydrodipicolinate synthase and dihydrodipicolinate reductase.	Corynebacterium glutamicum	99,539	16-Aug-93
		GB_PAT:E16749	2001	E16749	gDNA encoding dihydrodipicolinate synthase (DDPS).	Corynebacterium glutamicum	99,539	28-Jul-99
		GB_PAT:E14520	2001	E14520	DNA encoding Brevibacterium dihydrodipicolinic acid synthase.	Corynebacterium glutamicum	99,539	28-Jul-99
rxa00864	873	GB_BA1:BLDAPAB	3572	Z21502	B.lactofermentum dapA and dapB genes for dihydrodipicolinate synthase and dihydrodipicolinate reductase.	Corynebacterium glutamicum	99,885	16-Aug-93
		GB_BA1:CGDAPB	1902	X67737	C.glutamicum dapB gene for dihydrodipicolinate reductase.	Corynebacterium glutamicum	100,000	1-Apr-93
		GB_PAT:E14520	2001	E14520	DNA encoding Brevibacterium dihydrodipicolinic acid synthase.	Corynebacterium glutamicum	100,000	28-Jul-99
rxa00865	1026	GB_BA1:BLDAPAB	3572	Z21502	B.lactofermentum dapA and dapB genes for dihydrodipicolinate synthase and dihydrodipicolinate reductase.	Corynebacterium glutamicum	100,000	16-Aug-93
		GB_PAT:E16752	1411	E16752	gDNA encoding dihydrodipicolinate reductase (DDPR).	Corynebacterium glutamicum	99,805	28-Jul-99
		GB_PAT:AR038113	1411	AR038113	Sequence 18 from patent US 5804414.	Unknown.	99,805	29-Sep-99
rxa00867	650	GB_BA1:MTV002	56414	AL008967	Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.	Mycobacterium tuberculosis	39,179	17-Jun-98
		GB_BA1:MLCB22	40281	Z98741	Mycobacterium leprae cosmid B22.	Mycobacterium leprae	39,482	22-Aug-97
		GB_BA1:SAU19858	2838	U19858	Streptomyces antibioticus guanosine pentaphosphate synthetase (gpsI) gene, complete cds.	Streptomyces antibioticus	69,706	25-OCT-1996
rxa00873	779	GB_BA1:SCO001206	9184	AJ001206	Streptomyces coelicolor A3(2), glycogen metabolism cluster II.	Streptomyces coelicolor	63,415	29-MAR-1999
		GB_BA1:SCO001205	9589	AJ001205	Streptomyces coelicolor A3(2) glycogen metabolism clusterI.	Streptomyces coelicolor	61,617	29-MAR-1999

TABLE 4: ALIGNMENT RESULTS

		GB_BA1:D78198	2304	D78198	Pimelobacter sp. DNA for trehalose synthase, complete cds.	Pimelobacter sp.	60,594	5-Feb-99
rxa00884	1263	GB_BA1:MTCY253	41230	Z81368	Mycobacterium tuberculosis H37Rv complete genome; segment 106/162.	Mycobacterium tuberculosis	37,785	17-Jun-98
		GB_BA1:MSGY222	41156	AD000010	Mycobacterium tuberculosis sequence from clone y222.	Mycobacterium tuberculosis	38,006	03-DEC-1996
		GB_GSS15:AQ654600	468	AQ654600	Sheared DNA-1O14.TF Sheared DNA Trypanosoma brucei genomic clone Sheared DNA-1O14, genomic survey sequence.	Trypanosoma brucei	33,974	22-Jun-99
rxa00891	1102	GB_BA1:MTCI418B	11700	Z96071	Mycobacterium tuberculosis H37Rv complete genome; segment 7/162.	Mycobacterium tuberculosis	63,297	18-Jun-98
		GB_BA1:SCO001206	9184	AJ001206	Streptomyces coelicolor A3(2), glycogen metabolism cluster II.	Streptomyces coelicolor	61,965	29-MAR-1999
		GB_BA1:SCO001205	9589	AJ001205	Streptomyces coelicolor A3(2) glycogen metabolism clusterI.	Streptomyces coelicolor	61,727	29-MAR-1999
rxa00952	963	EM_PAT:E10963	3118	E10963	gDNA encoding tryptophan synthase.	Corynebacterium glutamicum	99,688	08-OCT-1997 (Rel. 52, Created)
		GB_BA1:BLTRP	7725	X04960	Brevibacterium lactofermentum tryptophan operon.	Corynebacterium glutamicum	98,847	10-Feb-99
		GB_PAT:E01688	7725	E01688	Genomic DNA of trp operon of prepibacterium latophelmentamn.	unidentified	98,428	29-Sep-97
rxa00954	644	GB_PAT:E01375	7726	E01375	DNA sequence of tryptophan operon.	Corynebacterium glutamicum	98,758	29-Sep-97
		GB_PAT:E01688	7725	E01688	Genomic DNA of trp operon of prepibacterium latophelmentamn.	unidentified	98,758	29-Sep-97
		GB_BA1:BLTRP	7725	X04960	Brevibacterium lactofermentum tryptophan operon.	Corynebacterium glutamicum	98,758	10-Feb-99
rxa00955	1545	GB_PAT:E01375	7726	E01375	DNA sequence of tryptophan operon.	Corynebacterium glutamicum	98,372	29-Sep-97
		GB_BA1:BLTRP	7725	X04960	Brevibacterium lactofermentum tryptophan operon.	Corynebacterium glutamicum	98,372	10-Feb-99
		GB_PAT:E01688	7725	E01688	Genomic DNA of trp operon of prepibacterium latophelmentamn.	unidentified	98,242	29-Sep-97
rxa00956	1237	EM_PAT:E10963	3118	E10963	gDNA encoding tryptophan synthase.	Corynebacterium glutamicum	98,949	08-OCT-1997 (Rel. 52, Created)
		GB_BA1:BLTRP	7725	X04960	Brevibacterium lactofermentum tryptophan operon.	Corynebacterium glutamicum	99,107	10-Feb-99
		GB_PAT:E01375	7726	E01375	DNA sequence of tryptophan operon.	Corynebacterium glutamicum	98,945	29-Sep-97
rxa00957	1677	GB_BA1:BLTRP	7725	X04960	Brevibacterium lactofermentum tryptophan operon.	Corynebacterium glutamicum	99,165	10-Feb-99
		GB_PAT:E01375	7726	E01375	DNA sequence of tryptophan operon.	Corynebacterium glutamicum	98,927	29-Sep-97
		GB_PAT:E01688	7725	E01688	Genomic DNA of trp operon of prepibacterium latophelmentamn.	unidentified	98,867	29-Sep-97
rxa00958	747	GB_BA1:BLTRP	7725	X04960	Brevibacterium lactofermentum tryptophan operon.	Corynebacterium glutamicum	98,792	10-Feb-99
		GB_PAT:E01375	7726	E01375	DNA sequence of tryptophan operon.	Corynebacterium glutamicum	98,792	29-Sep-97
		GB_PAT:E01688	7725	E01688	Genomic DNA of trp operon of prepibacterium latophelmentamn.	unidentified	98,658	29-Sep-97
rxa00970	1050	GB_BA1:CGHOMTHR	3685	Y00546	Corynebacterium glutamicum hom-thrB genes for homoserine dehydrogenase and homoserine kinase.	Corynebacterium glutamicum	99,905	12-Sep-93
		GB_PAT:I09077	3685	I09077	Sequence 1 from Patent WO 8809819.	Unknown.	99,810	02-DEC-1994
		GB_PAT:E01358	2615	E01358	DNA encoding for homoserine dehydrogenase(HDH)and homoserine kinase(HK).	Corynebacterium glutamicum	97,524	29-Sep-97
rxa00972	1458	GB_PAT:E16755	3579	E16755	gDNA encoding diaminopimelate decarboxylase (DDC) and arginyl-tRNA synthase.	Corynebacterium glutamicum	99,931	28-Jul-99
		GB_PAT:AR038110	3579	AR038110	Sequence 15 from patent US 5804414.	Unknown.	99,931	29-Sep-99
		GB_PAT:E14508	3579	E14508	DNA encoding Brevibacterium diaminopimelic acid decarboxylase and arginyl-tRNA synthase.	Corynebacterium glutamicum	99,931	28-Jul-99
rxa00981	753	GB_OV:GGA245664	512	AJ245664	Gallus gallus partial mRNA for ATP-citrate lyase (ACL gene).	Gallus gallus	37,538	28-Sep-99
		GB_PL2:AC007887	159434	AC007887	Genomic sequence for Arabidopsis thaliana BAC F15O4 from chromosome I, complete sequence.	Arabidopsis thaliana	37,600	04-OCT-1999

TABLE 4: ALIGNMENT RESULTS

		GB_GSS1:CNS00RNW	542	AL087338	Arabidopsis thaliana genome survey sequence T7 end of BAC F14D7 of IGF library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.	Arabidopsis thaliana	41,264	28-Jun-99
rxa00989	1644	GB_BA1:MTV008	63033	AL021246	Mycobacterium tuberculosis H37Rv complete genome; segment 108/162.	Mycobacterium tuberculosis	40,773	17-Jun-98
		GB_BA1:SCVALSFP	3619	Y13070	S.coelicolor valS, fpgs, ndk genes.	Streptomyces coelicolor	58,119	03-MAR-1998
		GB_BA1:MTV008	63033	AL021246	Mycobacterium tuberculosis H37Rv complete genome; segment 108/162.	Mycobacterium tuberculosis	38,167	17-Jun-98
rxa00997	705	GB_BA2:CGU31225	1817	U31225	Corynebacterium glutamicum L-proline:NADP+ 5-oxidoreductase (proC) gene, complete cds.	Corynebacterium glutamicum	40,841	2-Aug-96
		GB_HTG1:CEY39C12	282838	AL009026	Caenorhabditis elegans chromosome IV clone Y39C12, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	36,416	26-OCT-1999
		GB_IN1:CEB0001	39416	Z69634	Caenorhabditis elegans cosmid B0001, complete sequence.	Caenorhabditis elegans	36,416	2-Sep-99
rxa01019	1110	GB_HTG2:AC005052	144734	AC005052	Homo sapiens clone RG038K21, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	39,172	12-Jun-98
		GB_HTG2:AC005052	144734	AC005052	Homo sapiens clone RG038K21, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	Homo sapiens	39,172	12-Jun-98
		GB_GSS9:AQ171808	512	AQ171808	HS_3179_A1_G03_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3179 Col=5 Row=M, genomic survey sequence.	Homo sapiens	34,661	17-OCT-1998
rxa01026	1782	GB_BA1:SC1C2	42210	AL031124	Streptomyces coelicolor cosmid 1C2.	Streptomyces coelicolor	68,275	15-Jan-99
		GB_BA1:ATLEUCD	2982	X84647	A.teichomyceticus leuC and leuD genes.	Actinoplanes teichomyceticus	65,935	04-OCT-1995
		GB_BA1:MTV012	70287	AL021287	Mycobacterium tuberculosis H37Rv complete genome; segment 132/162.	Mycobacterium tuberculosis	40,454	23-Jun-99
rxa01027	1131	GB_BA1:MLCB637	44882	Z99263	Mycobacterium leprae cosmid B637.	Mycobacterium leprae	38,636	17-Sep-97
		GB_BA1:MTCY349	43523	Z83018	Mycobacterium tuberculosis H37Rv complete genome; segment 131/162.	Mycobacterium tuberculosis	51,989	17-Jun-98
		GB_BA1:SPUNGMUTX	1172	Z21702	S.pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase.	Streptococcus pneumoniae	38,088	15-Jun-94
rxa01073	954	GB_BA1:BACOUTB	1004	M15811	Bacillus subtilis outB gene encoding a sporulation protein, complete cds.	Bacillus subtilis	53,723	26-Apr-93
		GB_PR4:AC007938	167237	AC007938	Homo sapiens clone UWGC:djs201 from 7q31, complete sequence.	Homo sapiens	34,322	1-Jul-99
		GB_PL2:ATAC006282	92577	AC006282	Arabidopsis thaliana chromosome II BAC F13K3 genomic sequence, complete sequence.	Arabidopsis thaliana	36,181	13-MAR-1999
rxa01079	2226	GB_BA2:AF112535	4363	AF112535	Corynebacterium glutamicum putative glutaredoxin NrdH (nrdH), NrdI (nrdI), and ribonucleotide reductase alpha-chain (nrdE) genes, complete cds.	Corynebacterium glutamicum	99,820	5-Aug-99
		GB_BA1:CANRDFGEN	6054	Y09572	Corynebacterium ammoniagenes nrdH, nrdI, nrdE, nrdF genes.	Corynebacterium ammoniagenes	75,966	18-Apr-98
		GB_BA1:MTV012	70287	AL021287	Mycobacterium tuberculosis H37Rv complete genome; segment 132/162.	Mycobacterium tuberculosis	38,296	23-Jun-99
rxa01080	567	GB_BA2:AF112535	4363	AF112535	Corynebacterium glutamicum putative glutaredoxin NrdH (nrdH), NrdI (nrdI), and ribonucleotide reductase alpha-chain (nrdE) genes, complete cds.	Corynebacterium glutamicum	100,000	5-Aug-99
		GB_BA1:CANRDFGEN	6054	Y09572	Corynebacterium ammoniagenes nrdH, nrdI, nrdE, nrdF genes.	Corynebacterium ammoniagenes	65,511	18-Apr-98
		GB_BA1:STNRD	4894	X73226	S.typhimurium nrdEF operon.	Salmonella typhimurium	52,477	03-MAR-1997
rxa01087	999	GB_IN2:AF063412	1093	AF063412	Limnadia lenticularis elongation factor 1-alpha mRNA, partial cds.	Limnadia lenticularis	43,750	29-MAR-1999
		GB_PR3:HS24M15	134539	Z94055	Human DNA sequence from PAC 24M15 on chromosome 1. Contains tenascin-R (restrictin), EST.	Homo sapiens	37,475	23-Nov-99
		GB_IN2:ARU85702	1240	U85702	Anathix ralla elongation factor-1 alpha (EF-1a) gene, partial cds.	Anathix ralla	37,319	16-Jul-97

TABLE 4: ALIGNMENT RESULTS

rxa01095	857	GB_BA1:MTCY01B2	35938	Z95554	Mycobacterium tuberculosis H37Rv complete genome; segment 72/162.	Mycobacterium tuberculosis	43,243	17-Jun-98
		GB_HTG5:AC011632	175917	AC011632	Homo sapiens clone RP11-3N13, WORKING DRAFT SEQUENCE, 9 unordered pieces.	Homo sapiens	36,471	19-Nov-99
		GB_HTG5:AC011632	175917	AC011632	Homo sapiens clone RP11-3N13, WORKING DRAFT SEQUENCE, 9 unordered pieces.	Homo sapiens	36,836	19-Nov-99
rxa01097	477	GB_BA2:AF030405	774	AF030405	Corynebacterium glutamicum cyclase (hisF) gene, complete cds.	Corynebacterium glutamicum	100,000	13-Nov-97
		GB_BA2:AF030405	774	AF030405	Corynebacterium glutamicum cyclase (hisF) gene, complete cds.	Corynebacterium glutamicum	41,206	13-Nov-97
rxa01098	897	GB_BA2:AF030405	774	AF030405	Corynebacterium glutamicum cyclase (hisF) gene, complete cds.	Corynebacterium glutamicum	97,933	13-Nov-97
		GB_BA1:MSGY223	42061	AD000019	Mycobacterium tuberculosis sequence from clone y223.	Mycobacterium tuberculosis	40,972	10-DEC-1996
		GB_BA1:MLCB1610	40055	AL049913	Mycobacterium leprae cosmid B1610.	Mycobacterium leprae	61,366	27-Aug-99
rxa01100	861	GB_BA2:AF051846	738	AF051846	Corynebacterium glutamicum phosphoribosylformimino-5-amino-1-phosphoribosyl-4- imidazolecarboxamide isomerase (hisA) gene, complete cds.	Corynebacterium glutamicum	97,154	12-MAR-1998
		GB_BA2:AF060558	636	AF060558	Corynebacterium glutamicum glutamine amidotransferase (hisH) gene, complete cds.	Corynebacterium glutamicum	95,455	29-Apr-98
		GB_HTG1:HSDJ140A9	221755	AL109917	Homo sapiens chromosome 1 clone RP1-140A9, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	30,523	23-Nov-99
rxa01101	756	GB_BA2:AF060558	636	AF060558	Corynebacterium glutamicum glutamine amidotransferase (hisH) gene, complete cds.	Corynebacterium glutamicum	94,462	29-Apr-98
		GB_BA1:SC4G6	36917	AL096884	Streptomyces coelicolor cosmid 4G6.	Streptomyces coelicolor A3(2)	38,378	23-Jul-99
		GB_BA1:STMHISOPA	3981	M31628	S.coelicolor histidine biosynthesis operon encoding hisD, partial cds., and hisC, hisB, hisH, and hisA genes, complete cds.	Streptomyces coelicolor	60,053	26-Apr-93
rxa01104	729	GB_BA1:STMHISOPA	3981	M31628	S.coelicolor histidine biosynthesis operon encoding hisD, partial cds., and hisC, hisB, hisH, and hisA genes, complete cds.	Streptomyces coelicolor	58,333	26-Apr-93
		GB_BA1:SC4G6	36917	AL096884	Streptomyces coelicolor cosmid 4G6.	Streptomyces coelicolor A3(2)	39,045	23-Jul-99
		GB_BA1:MTCY336	32437	Z95586	Mycobacterium tuberculosis H37Rv complete genome; segment 70/162.	Mycobacterium tuberculosis	60,364	24-Jun-99
rxa01105	1221	GB_BA1:MTCY336	32437	Z95586	Mycobacterium tuberculosis H37Rv complete genome; segment 70/162.	Mycobacterium tuberculosis	60,931	24-Jun-99
		GB_BA1:MSGY223	42061	AD000019	Mycobacterium tuberculosis sequence from clone y223.	Mycobacterium tuberculosis	36,851	10-DEC-1996
		GB_BA1:MLCB1610	40055	AL049913	Mycobacterium leprae cosmid B1610.	Mycobacterium leprae	60,902	27-Aug-99
rxa01106	1449	GB_BA1:MSGY223	42061	AD000019	Mycobacterium tuberculosis sequence from clone y223.	Mycobacterium tuberculosis	37,233	10-DEC-1996
		GB_BA1:MSHISCD	2298	X65542	M.smegmatis genes hisD and hisC for histidinol dehydrogenase and histidinol-phosphate aminotransferase, respectively.	Mycobacterium smegmatis	60,111	30-Jun-93
		GB_BA1:MTCY336	32437	Z95586	Mycobacterium tuberculosis H37Rv complete genome; segment 70/162.	Mycobacterium tuberculosis	58,420	24-Jun-99
rxa01145	1137	GB_BA1:CORAIA	4705	L09232	Corynebacterium glutamicum acetohydroxy acid synthase (ilvB) and (ilvN) genes, and acetohydroxy acid isomeroreductase (ilvC) gene, complete cds.	Corynebacterium glutamicum	100,000	23-Feb-95
		GB_BA1:BRLILVCA	1364	D14551	Brevibacterium flavum ilvC gene for acetohydroxy acid isomeroreductase, complete cds.	Corynebacterium glutamicum	99,560	3-Feb-99
		GB_PAT:E08232	1017	E08232	DNA encoding acetohydroxy-acid isomeroreductase.	Corynebacterium glutamicum	99,803	29-Sep-97
rxa01162	1449	GB_PAT:A60299	2869	A60299	Sequence 18 from Patent WO9706261.	Aspergillus niger	38,675	06-MAR-1998
		GB_PR3:HS24E5	35506	Z82185	Human DNA sequence from Fosmid 24E5 on chromosome 22q11.2-qter contains parvalbumin, ESTs, STS.	Homo sapiens	36,204	23-Nov-99

TABLE 4: ALIGNMENT RESULTS

		GB_PR3:AC005265	43900	AC005265	Homo sapiens chromosome 19, cosmid F19750, complete sequence.	Homo sapiens	38,363	6-Jul-98
rxa01208	846	GB_HTG2:AC004965	323792	AC004965	Homo sapiens clone DJ1106H14, *** SEQUENCING IN PROGRESS ***, 42 unordered pieces.	Homo sapiens	36,058	12-Jun-98
		GB_HTG2:AC004965	323792	AC004965	Homo sapiens clone DJ1106H14, *** SEQUENCING IN PROGRESS ***, 42 unordered pieces.	Homo sapiens	36,058	12-Jun-98
		GB_PL2:TAU55859	2397	U55859	Triticum aestivum heat shock protein 80 mRNA, complete cds.	Triticum aestivum	37,269	1-Feb-99
rxa01209	1528	GB_HTG3:AC011469	113436	AC011469	Homo sapiens chromosome 19 clone CIT-HSPC_475D23, *** SEQUENCING IN PROGRESS ***, 31 unordered pieces.	Homo sapiens	40,000	07-OCT-1999
		GB_HTG3:AC011469	113436	AC011469	Homo sapiens chromosome 19 clone CIT-HSPC_475D23, *** SEQUENCING IN PROGRESS ***, 31 unordered pieces.	Homo sapiens	40,000	07-OCT-1999
		GB_PL1:AB010077	77380	AB010077	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MYH19, complete sequence.	Arabidopsis thaliana	36,803	20-Nov-99
rxa01215	1098	GB_BA1:MTCY10G2	38970	Z92539	Mycobacterium tuberculosis H37Rv complete genome; segment 47/162.	Mycobacterium tuberculosis	37,047	17-Jun-98
		GB_IN1:LEIPRPP	1887	M76553	Leishmania donovani phosphoribosylpyrophosphate synthetase gene, complete cds.	Leishmania donovani	50,738	7-Jun-93
		GB_HTG2:HSJ799D16	130149	AL050344	Homo sapiens chromosome 1 clone RP4-799D16 map p34.3-36.1, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	38,135	29-Nov-99
rxa01239	2556	GB_BA1:MTCY48	35377	Z74020	Mycobacterium tuberculosis H37Rv complete genome; segment 69/162.	Mycobacterium tuberculosis	38,139	17-Jun-98
		GB_PR2:AB029032	6377	AB029032	Homo sapiens mRNA for KIAA1109 protein, partial cds.	Homo sapiens	39,394	4-Aug-99
		GB_GSS9:AQ107201	355	AQ107201	HS_3098_A1_C03_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3098 Col=5 Row=E, genomic survey sequence.	Homo sapiens	41,408	28-Aug-98
rxa01253	873	GB_PL2:F5O8	99923	AC005990	Arabidopsis thaliana chromosome 1 BAC F5O8 sequence, complete sequence.	Arabidopsis thaliana	36,118	23-DEC-1998
		GB_PL2:F5O8	99923	AC005990	Arabidopsis thaliana chromosome 1 BAC F5O8 sequence, complete sequence.	Arabidopsis thaliana	35,574	23-DEC-1998
		GB_IN1:CELC06G1	31205	U41014	Caenorhabditis elegans cosmid C06G1.	Caenorhabditis elegans	38,560	30-Nov-95
rxa01321	1044	GB_GSS14:AQ518843	441	AQ518843	HS_5106_A1_D10_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=682 Col=19 Row=G, genomic survey sequence.	Homo sapiens	41,121	05-MAY-1999
		GB_HTG2:AC007473	194859	AC007473	Drosophila melanogaster chromosome 2 clone BACR38D12 (D590) RPCI-98 38.D.12 map 48A-48B strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 60 unordered pieces.	Drosophila melanogaster	40,634	2-Aug-99
		GB_HTG4:AC011696	115847	AC011696	Drosophila melanogaster chromosome 2 clone BACR35F01 (D1156) RPCI-98 35.F.1 map 48A-48C strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 108 unordered pieces.	Drosophila melanogaster	38,290	26-OCT-1999
rxa01352	706	GB_PL2:ATAC005167	83260	AC005167	Arabidopsis thaliana chromosome II BAC F12A24 genomic sequence, complete sequence.	Arabidopsis thaliana	34,311	15-OCT-1998
		GB_PL2:ATAC005825	97380	AC005825	Arabidopsis thaliana chromosome II BAC T24I21 genomic sequence, complete sequence.	Arabidopsis thaliana	34,311	12-Apr-99
		GB_HTG3:AC011150	127222	AC011150	Homo sapiens clone 4_K_17, LOW-PASS SEQUENCE SAMPLING.	Homo sapiens	37,722	01-OCT-1999
rxa01360	259	GB_EST32:AI725583	728	AI725583	BNLGHi12371 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (U86081) root hair defective 3 [Arabidopsis thaliana], mRNA sequence.	Gossypium hirsutum	38,492	11-Jun-99
		GB_PR2:HS227P17	82951	Z81007	Human DNA sequence from PAC 227P17, between markers DXS6791 andDXS8038 on chromosome X contains CpG island, EST.	Homo sapiens	39,738	23-Nov-99

TABLE 4: ALIGNMENT RESULTS

		GB_EST34:AV171099	173	AV171099	AV171099 Mus musculus head C57BL/6J 14, 17 day embryo Mus musculus cDNA clone 3200002M11, mRNA sequence.	Mus musculus	46,237	6-Jul-99
rxa01361	629	GB_RO:AB008915S1	530	AB008915	Mus musculus mGpi1 gene, exon 1.	Mus musculus	45,574	28-Sep-99
		GB_EST22:AI050532	293	AI050532	uc83d10.y1 Sugano mouse kidney mkia Mus musculus cDNA clone IMAGE:1432243 5' similar to TR:O35120 O35120 MGPI1P. ;, mRNA sequence.	Mus musculus	44,097	9-Jul-98
		GB_RO:AB008895	3062	AB008895	Mus musculus mRNA for mGpi1p, complete cds.	Mus musculus	41,316	23-Nov-97
rxa01381	944	GB_PL1:AB005237	87835	AB005237	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MJJ3,complete sequence.	Arabidopsis thaliana	36,606	20-Nov-99
		GB_GSS5:AQ766840	491	AQ766840	HS_2026_A2_C09_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2026 Col=18 Row=E, genomic survey sequence.	Homo sapiens	37,916	28-Jul-99
		GB_BA1:MTV043	68848	AL022004	Mycobacterium tuberculosis H37Rv complete genome; segment 40/162.	Mycobacterium tuberculosis	37,419	24-Jun-99
rxa01393	993	GB_BA1:CGLYSEG	2374	X96471	C.glutamicum lysE and lysG genes.	Corynebacterium glutamicum	34,831	24-Feb-97
		GB_BA1:SC5A7	40337	AL031107	Streptomyces coelicolor cosmid 5A7.	Streptomyces coelicolor	35,138	27-Jul-98
		GB_PR3:AC004054	112184	AC004054	Homo sapiens chromosome 4 clone B220G8 map 4q21, complete sequence.	Homo sapiens	37,277	9-Jul-98
rxa01394	822	GB_BA1:CGLYSEG	2374	X96471	C.glutamicum lysE and lysG genes.	Corynebacterium glutamicum	100,000	24-Feb-97
		GB_GSS5:AQ769223	500	AQ769223	HS_3155_B2_G10_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3155 Col=20 Row=N, genomic survey sequence.	Homo sapiens	38,400	28-Jul-99
		GB_BA1:CGLYSEG	2374	X96471	C.glutamicum lysE and lysG genes.	Corynebacterium glutamicum	33,665	24-Feb-97
rxa01416	630	GB_BA1:SC3C3	31382	AL031231	Streptomyces coelicolor cosmid 3C3.	Streptomyces coelicolor	62,726	10-Aug-98
		GB_BA1:MLCB22	40281	Z98741	Mycobacterium leprae cosmid B22.	Mycobacterium leprae	39,159	22-Aug-97
		GB_BA1:MTV002	56414	AL008967	Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.	Mycobacterium tuberculosis	37,340	17-Jun-98
rxa01442	1347	GB_BA1:D90827	18886	D90827	E.coli genomic DNA, Kohara clone #336(41.2-41.6 min.).	Escherichia coli	58,517	21-MAR-1997
		GB_BA1:D90828	14590	D90828	E.coli genomic DNA, Kohara clone #336gap(41.6-41.9 min.).	Escherichia coli	56,151	21-MAR-1997
		GB_BA2:AE000279	10855	AE000279	Escherichia coli K-12 MG1655 section 169 of 400 of the complete genome.	Escherichia coli	56,021	12-Nov-98
rxa01446	1413	GB_BA1:SCH10	39524	AL049754	Streptomyces coelicolor cosmid H10.	Streptomyces coelicolor	39,037	04-MAY-1999
		GB_BA1:MTY13E10	35019	Z95324	Mycobacterium tuberculosis H37Rv complete genome; segment 18/162.	Mycobacterium tuberculosis	40,130	17-Jun-98
		GB_BA1:MLCB4	36310	AL023514	Mycobacterium leprae cosmid B4.	Mycobacterium leprae	37,752	27-Aug-99
rxa01483	1395	GB_BA1:MTCY98	31225	Z83860	Mycobacterium tuberculosis H37Rv complete genome; segment 103/162.	Mycobacterium tuberculosis	39,057	17-Jun-98
		GB_BA1:MSGB1229CS	30670	L78812	Mycobacterium leprae cosmid B1229 DNA sequence.	Mycobacterium leprae	54,382	15-Jun-96
		GB_BA2:AF027507	5168	AF027507	Mycobacterium smegmatis dGTPase (dgt), and primase (dnaG) genes, complete cds; tRNA-Asn gene, complete sequence.	Mycobacterium smegmatis	52,941	16-Jan-98
rxa01486	757	GB_BA1:MTV002	56414	AL008967	Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.	Mycobacterium tuberculosis	40,941	17-Jun-98
		GB_BA1:MLCB22	40281	Z98741	Mycobacterium leprae cosmid B22.	Mycobacterium leprae	38,451	22-Aug-97
		GB_BA1:SC3C3	31382	AL031231	Streptomyces coelicolor cosmid 3C3.	Streptomyces coelicolor	61,194	10-Aug-98
rxa01489	1146	GB_BA1:CORFADS	1547	D37967	Corynebacterium ammoniagenes gene for FAD synthetase, complete cds.	Corynebacterium ammoniagenes	58,021	8-Feb-99
		GB_BA1:MLCB22	40281	Z98741	Mycobacterium leprae cosmid B22.	Mycobacterium leprae	38,414	22-Aug-97
		GB_BA1:SC10A7	39739	AL078618	Streptomyces coelicolor cosmid 10A7.	Streptomyces coelicolor	36,930	9-Jun-99
rxa01491	774	GB_BA1:MTV002	56414	AL008967	Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.	Mycobacterium tuberculosis	37,062	17-Jun-98

TABLE 4: ALIGNMENT RESULTS

		GB_EST13:AA356956	255	AA356956	EST65614 Jurkat T-cells III Homo sapiens cDNA 5' end, mRNA sequence.	Homo sapiens	37,647	21-Apr-97
		GB_OV:OMDNAPROI	7327	X92380	O.mossambicus prolactin I gene.	Tilapia mossambica	38,289	19-OCT-1995
rxa01508	1662	GB_IN1:CEF28C12	14653	Z93380	Caenorhabditis elegans cosmid F28C12, complete sequence.	Caenorhabditis elegans	37,984	23-Nov-98
		GB_IN1:CEF28C12	14653	Z93380	Caenorhabditis elegans cosmid F28C12, complete sequence.	Caenorhabditis elegans	38,469	23-Nov-98
rxa01512	723	GB_BA1:SCE9	37730	AL049841	Streptomyces coelicolor cosmid E9.	Streptomyces coelicolor	39,021	19-MAY-1999
		GB_BA1:MAU88875	840	U88875	Mycobacterium avium hypoxanthine-guanine phosphoribosyl transferase gene, complete cds.	Mycobacterium avium	57,521	05-MAR-1997
		GB_BA1:MTY15C10	33050	Z95436	Mycobacterium tuberculosis H37Rv complete genome; segment 154/162.	Mycobacterium tuberculosis	40,086	17-Jun-98
rxa01514	711	GB_BA1:MTCY7H7B	24244	Z95557	Mycobacterium tuberculosis H37Rv complete genome; segment 153/162.	Mycobacterium tuberculosis	43,343	18-Jun-98
		GB_BA1:MLCB2548	38916	AL023093	Mycobacterium leprae cosmid B2548.	Mycobacterium leprae	38,177	27-Aug-99
		GB_PL1:EGGTPCHI	242	Z49757	E.gracilis mRNA for GTP cyclohydrolase I (core region).	Euglena gracilis	64,876	20-OCT-1995
rxa01515	975	GB_BA1:ECOUW93	338534	U14003	Escherichia coli K-12 chromosomal region from 92.8 to 00.1 minutes.	Escherichia coli	38,943	17-Apr-96
		GB_BA1:ECOUW93	338534	U14003	Escherichia coli K-12 chromosomal region from 92.8 to 00.1 minutes.	Escherichia coli	37,500	17-Apr-96
		GB_BA1:MTCY49	39430	Z73966	Mycobacterium tuberculosis H37Rv complete genome; segment 93/162.	Mycobacterium tuberculosis	38,010	24-Jun-99
rxa01516	513	GB_IN1:DME238847	5419	AJ238847	Drosophila melanogaster mRNA for drosophila dodeca-satellite protein 1 (DDP-1).	Drosophila melanogaster	36,346	13-Aug-99
		GB_HTG3:AC009210	103814	AC009210	Drosophila melanogaster chromosome 2 clone BACR01I06 (D1054) RPCI-98 01.I.6 map 55D-55D strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 86 unordered pieces.	Drosophila melanogaster	37,897	20-Aug-99
		GB_IN2:AF132179	4842	AF132179	Drosophila melanogaster clone LD21677 unknown mRNA.	Drosophila melanogaster	36,149	3-Jun-99
rxa01517	600	GB_PL2:F6H8	82596	AF178045	Arabidopsis thaliana BAC F6H8.	Arabidopsis thaliana	35,846	19-Aug-99
		GB_PL2:AF038831	647	AF038831	Sorosporium saponariae internal transcribed spacer 1, 5.8S ribosomal RNA gene; and internal transcribed spacer 2, complete sequence.	Sorosporium saponariae	40,566	13-Apr-99
		GB_PL2:ATAC005957	108355	AC005957	Arabidopsis thaliana chromosome II BAC T15J14 genomic sequence, complete sequence.	Arabidopsis thaliana	38,095	7-Jan-99
rxa01521	921	GB_BA1:ANANIFBH	5936	J05111	Anabaena sp. (clone AnH20.1) nitrogen fixation operon nifB, fdxN, nifS, nifU, and nifH genes, complete cds.	Anabaena sp.	38,206	26-Apr-93
		GB_PR2:AC002461	197273	AC002461	Human BAC clone RG204I16 from 7q31, complete sequence.	Homo sapiens	36,623	20-Aug-97
		GB_PR2:AC002461	197273	AC002461	Human BAC clone RG204I16 from 7q31, complete sequence.	Homo sapiens	34,719	20-Aug-97
rxa01528	651	GB_RO:MM437P9	165901	AL049866	Mus musculus chromosome X, clone 437P9.	Mus musculus	37,500	29-Jun-99
		GB_PR3:AC005740	186780	AC005740	Homo sapiens chromosome 5p, BAC clone 50g21 (LBNL H154), complete sequence.	Homo sapiens	37,031	01-OCT-1998
		GB_PR3:AC005740	186780	AC005740	Homo sapiens chromosome 5p, BAC clone 50g21 (LBNL H154), complete sequence.	Homo sapiens	38,035	01-OCT-1998
rxa01551	1998	GB_BA1:MTCY22G10	35420	Z84724	Mycobacterium tuberculosis H37Rv complete genome; segment 21/162.	Mycobacterium tuberculosis	38,371	17-Jun-98
		GB_BA2:ECOUW89	176195	U00006	E. coli chromosomal region from 89.2 to 92.8 minutes.	Escherichia coli	38,064	17-DEC-1993
		GB_BA1:SCQ11	15441	AL096823	Streptomyces coelicolor cosmid Q11.	Streptomyces coelicolor	60,775	8-Jul-99
rxa01561	1053	GB_IN1:CEY62H9A	47396	AL032630	Caenorhabditis elegans cosmid Y62H9A, complete sequence.	Caenorhabditis elegans	38,514	2-Sep-99
		GB_PR4:HSU51003	3202	U51003	Homo sapiens DLX-2 (DLX-2) gene, complete cds.	Homo sapiens	37,730	07-DEC-1999
		GB_OM:PIGDAO1	395	M18444	Pig D-amino acid oxidase (DAO) gene, exon 1.	Sus scrofa	39,340	27-Apr-93
rxa01599	1785	GB_BA1:MTCI125	37432	Z98268	Mycobacterium tuberculosis H37Rv complete genome; segment 76/162.	Mycobacterium tuberculosis	63,300	17-Jun-98
		GB_BA1:U00021	39193	U00021	Mycobacterium leprae cosmid L247.	Mycobacterium leprae	36,756	29-Sep-94
		GB_BA1:MLCB1351	38936	Z95117	Mycobacterium leprae cosmid B1351.	Mycobacterium leprae	36,756	24-Jun-97

TABLE 4: ALIGNMENT RESULTS

rxa01617	795	GB_PR2:HSMTM0	217657	AL034384	Human chromosome Xq28, cosmid clones 7H3, 14D7, C1230, 11E7, F1096, A12197, 12G8, A09100; complete sequence bases 1..217657.	Homo sapiens	40,811	5-Jul-99
		GB_PR2:HS13D10	153147	AL021407	Homo sapiens DNA sequence from PAC 13D10 on chromosome 6p22.3-23. Contains CpG island.	Homo sapiens	38,768	23-Nov-99
		GB_PR2:HSMTM0	217657	AL034384	Human chromosome Xq28, cosmid clones 7H3, 14D7, C1230, 11E7, F1096, A12197, 12G8, A09100; complete sequence bases 1..217657.	Homo sapiens	39,018	5-Jul-99
rxa01657	723	GB_BA1:MTCY1A10	25949	Z95387	Mycobacterium tuberculosis H37Rv complete genome; segment 117/162.	Mycobacterium tuberculosis	40,656	17-Jun-98
		GB_EST6:D79278	392	D79278	HUM213D06B Human aorta polyA+ (TFujiwara) Homo sapiens cDNA clone GEN-213D06 5', mRNA sequence.	Homo sapiens	44,262	9-Feb-96
		GB_BA2:AF129925	10243	AF129925	Thiobacillus ferrooxidans carboxysome operon, complete cds.	Thiobacillus ferrooxidans	40,709	17-MAY-1999
rxa01660	675	GB_BA1:MTV013	11364	AL021309	Mycobacterium tuberculosis H37Rv complete genome; segment 134/162.	Mycobacterium tuberculosis	40,986	17-Jun-98
		GB_RO:MMFV1	6480	X97719	M.musculus retrovirus restriction gene Fv1.	Mus musculus	35,364	29-Aug-96
		GB_PAT:A67508	6480	A67508	Sequence 1 from Patent WO9743410.	Mus musculus	35,364	05-MAY-1999
rxa01678	651	GB_VI:TVU95309	600	U95309	Tula virus O64 nucleocapsid protein gene, partial cds.	Tula virus	41,894	28-OCT-1997
		GB_VI:TVU95303	600	U95303	Tula virus O52 nucleocapsid protein gene, partial cds.	Tula virus	41,712	28-OCT-1997
		GB_VI:TVU95302	600	U95302	Tula virus O24 nucleocapsid protein gene, partial cds.	Tula virus	39,576	28-OCT-1997
rxa01679	1359	GB_EST5:H91843	362	H91843	ys81e01.s1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:221208 3' similar to gb:X63749_rna1 GUANINE NUCLEOTIDE-BINDING PROTEIN G(T), ALPHA-1 (HUMAN);, mRNA sequence.	Homo sapiens	39,157	29-Nov-95
		GB_STS:G26925	362	G26925	human STS SHGC-30023, sequence tagged site.	Homo sapiens	39,157	14-Jun-96
		GB_PL2:AF139451	1202	AF139451	Gossypium robinsonii CelA2 pseudogene, partial sequence.	Gossypium robinsonii	38,910	1-Jun-99
rxa01690	1224	GB_BA1:SC1C2	42210	AL031124	Streptomyces coelicolor cosmid 1C2.	Streptomyces coelicolor	60,644	15-Jan-99
		GB_EST22:AI064232	493	AI064232	GH04563.5prime GH Drosophila melanogaster head pOT2 Drosophila melanogaster cDNA clone GH04563 5prime, mRNA sequence.	Drosophila melanogaster	38,037	24-Nov-98
		GB_IN2:AF117896	1020	AF117896	Drosophila melanogaster neuropeptide F (npf) gene, complete cds.	Drosophila melanogaster	36,122	2-Jul-99
rxa01692	873	GB_BA2:AF067123	1034	AF067123	Lactobacillus reuteri cobalamin biosynthesis protein J (cbiJ) gene, partial cds; and uroporphyrin-III C-methyltransferase (sumT) gene, complete cds.	Lactobacillus reuteri	48,079	3-Jun-98
		GB_RO:RATNFHPEP	3085	M37227	Rat heavy neurofilament (NF-H) polypeptide, partial cds.	Rattus norvegicus	37,093	27-Apr-93
		GB_RO:RSNFH	3085	X13804	Rat mRNA for heavy neurofilament polypeptide NF-H C-terminus.	Rattus sp.	37,093	14-Jul-95
rxa01698	1353	GB_BA2:AF124600	4115	AF124600	Corynebacterium glutamicum chorismate synthase (aroC), shikimate kinase (aroK), and 3-dehydroquinate synthase (aroB) genes, complete cds; and putative cytoplasmic peptidase (pepQ) gene, partial cds.	Corynebacterium glutamicum	100,000	04-MAY-1999
		GB_BA1:MTCY159	33818	Z83863	Mycobacterium tuberculosis H37Rv complete genome; segment 111/162.	Mycobacterium tuberculosis	36,323	17-Jun-98
		GB_BA1:MSGB937CS	38914	L78820	Mycobacterium leprae cosmid B937 DNA sequence.	Mycobacterium leprae	62,780	15-Jun-96
rxa01699	693	GB_BA2:AF124600	4115	AF124600	Corynebacterium glutamicum chorismate synthase (aroC), shikimate kinase (aroK), and 3-dehydroquinate synthase (aroB) genes, complete cds; and putative cytoplasmic peptidase (pepQ) gene, partial cds.	Corynebacterium glutamicum	100,000	04-MAY-1999
		GB_BA2:AF016585	41097	AF016585	Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi) gene, partial cds; polyketide synthase modules 1 through 7 (nidA) genes, complete cds; and N-methyltransferase homolog gene, partial cds.	Streptomyces caelestis	40,260	07-DEC-1997
		GB_EST9:C19712	399	C19712	C19712 Rice panicle at ripening stage Oryza sativa cDNA clone E10821_1A, mRNA sequence.	Oryza sativa	45,425	24-OCT-1996
rxa01712	805	GB_EST21:AA952466	278	AA952466	TENS1404 T. cruzi epimastigote normalized cDNA Library Trypanosoma cruzi cDNA clone 1404 5', mRNA sequence.	Trypanosoma cruzi	40,876	29-OCT-1998

TABLE 4: ALIGNMENT RESULTS

		GB_EST21:AA952466	278	AA952466	TENS1404 T. cruzi epimastigote normalized cDNA Library Trypanosoma cruzi cDNA clone 1404 5', mRNA sequence.	Trypanosoma cruzi	41,367	29-OCT-1998
rxa01719	684	GB_HTG1:HSDJ534K7	154416	AL109925	Homo sapiens chromosome 1 clone RP4-534K7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	35,651	23-Nov-99
		GB_HTG1:HSDJ534K7	154416	AL109925	Homo sapiens chromosome 1 clone RP4-534K7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	35,651	23-Nov-99
		GB_EST27:AI447108	431	AI447108	mq91e08.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone IMAGE:586118 3', mRNA sequence.	Mus musculus	39,671	09-MAR-1999
rxa01720	1332	GB_PR4:AC006322	179640	AC006322	Homo sapiens PAC clone DJ1060B11 from 7q11.23-q21.1, complete sequence.	Homo sapiens	35,817	18-MAR-1999
		GB_PL2:TM018A10	106184	AF013294	Arabidopsis thaliana BAC TM018A10.	Arabidopsis thaliana	35,698	12-Jul-97
		GB_PR4:AC006322	179640	AC006322	Homo sapiens PAC clone DJ1060B11 from 7q11.23-q21.1, complete sequence.	Homo sapiens	37,243	18-MAR-1999
rxa01746	876	GB_EST3:R46227	443	R46227	yg52a03.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:36000 3', mRNA sequence.	Homo sapiens	42,812	22-MAY-1995
		GB_EST3:R46227	443	R46227	yg52a03.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:36000 3', mRNA sequence.	Homo sapiens	42,655	22-MAY-1995
rxa01747	1167	GB_BA1:MTCY190	34150	Z70283	Mycobacterium tuberculosis H37Rv complete genome; segment 98/162.	Mycobacterium tuberculosis	59,294	17-Jun-98
		GB_BA1:MLCB22	40281	Z98741	Mycobacterium leprae cosmid B22.	Mycobacterium leprae	57,584	22-Aug-97
		GB_BA1:SC5F7	40024	AL096872	Streptomyces coelicolor cosmid 5F7.	Streptomyces coelicolor A3(2)	61,810	22-Jul-99
rxa01757	924	GB_EST21:AA918454	416	AA918454	om38c02.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1543298 3' similar to WP:F28F8.3 CE09757 SMALL NUCLEAR RIBONUCLEOPROTEIN E ;, mRNA sequence.	Homo sapiens	39,655	23-Jun-98
		GB_EST4:H34042	345	H34042	EST110563 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA clone RPNBI81 5' end, mRNA sequence.	Rattus sp.	35,942	2-Apr-98
		GB_EST20:AA899038	450	AA899038	NCP6G8T7 Perithecial Neurospora crassa cDNA clone NP6G8 3' end, mRNA sequence.	Neurospora crassa	40,000	12-Apr-98
rxa01807	915	GB_BA1:AP000063	185300	AP000063	Aeropyrum pernix genomic DNA, section 6/7.	Aeropyrum pernix	40,067	22-Jun-99
		GB_HTG4:AC010694	115857	AC010694	Drosophila melanogaster clone RPCI98-6H2, *** SEQUENCING IN PROGRESS ***, 75 unordered pieces.	Drosophila melanogaster	35,450	16-OCT-1999
		GB_HTG4:AC010694	115857	AC010694	Drosophila melanogaster clone RPCI98-6H2, *** SEQUENCING IN PROGRESS ***, 75 unordered pieces.	Drosophila melanogaster	35,450	16-OCT-1999
rxa01821	401	GB_BA1:CGL007732	4460	AJ007732	Corynebacterium glutamicum 3' ppc gene, secG gene, amt gene, ocd gene and 5' soxA gene.	Corynebacterium glutamicum	100,000	7-Jan-99
		GB_RO:RATALGL	7601	M24108	Rattus norvegicus (clone A2U42) alpha2u globulin gene, exons 1-7.	Rattus norvegicus	38,692	15-DEC-1994
		GB_OV:APIGY2	1381	X78272	Anas platyrhynchos (Super M) IgY upsilon heavy chain gene, exon 2.	Anas platyrhynchos	36,962	15-Feb-99
rxa01835	654	GB_EST30:AI629479	353	AI629479	486101D10.x1 486 - leaf primordia cDNA library from Hake lab Zea mays cDNA, mRNA sequence.	Zea mays	38,109	26-Apr-99
		GB_STS:G48245	515	G48245	SHGC-62915 Human Homo sapiens STS genomic, sequence tagged site.	Homo sapiens	37,021	26-MAR-1999
		GB_GSS3:B49052	515	B49052	RPCI11-4I12.TV RPCI-11 Homo sapiens genomic clone RPCI-11-4I12, genomic survey sequence.	Homo sapiens	37,021	8-Apr-99

TABLE 4: ALIGNMENT RESULTS

rxa01850	1470	GB_BA2:ECOUW67_0	110000	U18997	Escherichia coli K-12 chromosomal region from 67.4 to 76.0 minutes.	Escherichia coli	37,196	U18997
		GB_BA2:AE000392	10345	AE000392	Escherichia coli K-12 MG1655 section 282 of 400 of the complete genome.	Escherichia coli	38,021	12-Nov-98
		GB_BA2:U32715	13136	U32715	Haemophilus influenzae Rd section 30 of 163 of the complete genome.	Haemophilus influenzae Rd	39,860	29-MAY-1998
rxa01878	1002	GB_HTG1:CEY64F11	177748	Z99776	Caenorhabditis elegans chromosome IV clone Y64F11, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	37,564	14-OCT-1998
		GB_HTG1:CEY64F11	177748	Z99776	Caenorhabditis elegans chromosome IV clone Y64F11, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	37,564	14-OCT-1998
		GB_HTG1:CEY64F11	177748	Z99776	Caenorhabditis elegans chromosome IV clone Y64F11, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Caenorhabditis elegans	37,576	14-OCT-1998
rxa01892	852	GB_BA1:MTCY274	39991	Z74024	Mycobacterium tuberculosis H37Rv complete genome; segment 126/162.	Mycobacterium tuberculosis	35,910	19-Jun-98
		GB_BA1:MLCB250	40603	Z97369	Mycobacterium leprae cosmid B250.	Mycobacterium leprae	64,260	27-Aug-99
		GB_BA1:MSGB1529CS	36985	L78824	Mycobacterium leprae cosmid B1529 DNA sequence.	Mycobacterium leprae	64,260	15-Jun-96
rxa01894	978	GB_BA1:MTCY274	39991	Z74024	Mycobacterium tuberculosis H37Rv complete genome; segment 126/162.	Mycobacterium tuberculosis	37,229	19-Jun-98
		GB_IN1:CELF46H5	38886	U41543	Caenorhabditis elegans cosmid F46H5.	Caenorhabditis elegans	38,525	29-Nov-96
		GB_HTG3:AC009204	115633	AC009204	Drosophila melanogaster chromosome 2 clone BACR03E19 (D1033) RPCI-98 03.E.19 map 36E-37C strain y; cn bw sp. *** SEQUENCING IN PROGRESS ***, 94 unordered pieces.	Drosophila melanogaster	31,579	18-Aug-99
rxa01920	1125	GB_BA2:AF112536	1798	AF112536	Corynebacterium glutamicum ribonucleotide reductase beta-chain (nrdF) gene, complete cds.	Corynebacterium glutamicum	99,733	5-Aug-99
		GB_BA1:CANRDFGEN	6054	Y09572	Corynebacterium ammoniagenes nrdH, nrdI, nrdE, nrdF genes.	Corynebacterium ammoniagenes	70,321	18-Apr-98
		GB_BA2:AF050168	1228	AF050168	Corynebacterium ammoniagenes ribonucleoside diphosphate reductase small subunit (nrdF) gene, complete cds.	Corynebacterium ammoniagenes	72,082	23-Apr-98
rxa01928	960	GB_BA1:CGPAN	2164	X96580	C.glutamicum panB, panC & xylB genes.	Corynebacterium glutamicum	100,000	11-MAY-1999
		GB_PL1:AP000423	154478	AP000423	Arabidopsis thaliana chloroplast genomic DNA, complete sequence, strain:Columbia.	Chloroplast Arabidopsis thaliana	35,917	15-Sep-99
		GB_PL1:AP000423	154478	AP000423	Arabidopsis thaliana chloroplast genomic DNA, complete sequence, strain:Columbia.	Chloroplast Arabidopsis thaliana	33,925	15-Sep-99
rxa01929	936	GB_BA1:CGPAN	2164	X96580	C.glutamicum panB, panC & xylB genes.	Corynebacterium glutamicum	100,000	11-MAY-1999
		GB_BA1:XCU33548	8429	U33548	Xanthomonas campestris hrpB pathogenicity locus proteins HrpB1, HrpB2, HrpB3, HrpB4, HrpB5, HrpB6, HrpB7, HrpB8, HrpA1, and ORF62 genes, complete cds.	Xanthomonas campestris pv. vesicatoria	38,749	19-Sep-96
		GB_BA1:XANHRPB6A	1329	M99174	Xanthomonas campestris hrpB6 gene, complete cds.	Xanthomonas campestris	39,305	14-Sep-93
rxa01940	1059	GB_IN2:CFU43371	1060	U43371	Crithidia fasciculata inosine-uridine preferring nucleoside hydrolase (IUNH) gene, complete cds.	Crithidia fasciculata	61,417	18-Jun-96
		GB_BA2:AE001467	11601	AE001467	Helicobacter pylori, strain J99 section 28 of 132 of the complete genome.	Helicobacter pylori J99	38,560	20-Jan-99
		GB_RO:AF175967	3492	AF175967	Homo sapiens Leman coiled-coil protein (LCCP) mRNA, complete cds.	Mus musculus	40,275	26-Sep-99
rxa02022	1230	GB_BA1:CGDAPE	1966	X81379	C.glutamicum dapE gene and orf2.	Corynebacterium glutamicum	100,000	8-Aug-95
		GB_BA1:CGDNAAROP	2612	X85965	C.glutamicum ORF3 and aroP gene.	Corynebacterium glutamicum	38,889	30-Nov-97
		GB_BA1:APU47055	6469	U47055	Anabaena PCC7120 nitrogen fixation proteins (nifE, nifN, nifX, nifW) genes, complete cds, and nitrogenase (nifK) and hesA genes, partial cds.	Anabaena PCC7120	36,647	17-Feb-96
rxa02024	859	GB_BA1:MTCI364	29540	Z93777	Mycobacterium tuberculosis H37Rv complete genome; segment 52/162.	Mycobacterium tuberculosis	59,415	17-Jun-98

TABLE 4: ALIGNMENT RESULTS

		GB_BA1:MSGB1912CS	38503	L01536	M. leprae genomic dna sequence, cosmid b1912.	Mycobacterium leprae	57,093	14-Jun-96
		GB_BA1:MLU15180	38675	U15180	Mycobacterium leprae cosmid B1756.	Mycobacterium leprae	57,210	09-MAR-1995
rxa02027								
rxa02031								
rxa02072	1464	GB_BA1:CGGDHA	2037	X72855	C.glutamicum GDHA gene.	Corynebacterium glutamicum	99,317	24-MAY-1993
		GB_BA1:CGGDH	2037	X59404	Corynebacterium glutamicum, gdh gen for glutamate dehydrogenase.	Corynebacterium glutamicum	94,387	30-Jul-99
		GB_BA1:PAE18494	1628	Y18494	Pseudomonas aeruginosa gdhA gene, strain PAC1.	Pseudomonas aeruginosa	62,247	6-Feb-99
rxa02085	2358	GB_BA1:MTCY22G8	22550	Z95585	Mycobacterium tuberculosis H37Rv complete genome; segment 49/162.	Mycobacterium tuberculosis	38,442	17-Jun-98
		GB_BA1:MLCB33	42224	Z94723	Mycobacterium leprae cosmid B33.	Mycobacterium leprae	56,486	24-Jun-97
		GB_BA1:ECOUW85	91414	M87049	E. coli genomic sequence of the region from 84.5 to 86.5 minutes.	Escherichia coli	52,127	29-MAY-1995
rxa02093	927	GB_EST14:AA448146	452	AA448146	zw82h01.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:782737 5', mRNA sequence.	Homo sapiens	34,163	4-Jun-97
		GB_EST17:AA641937	444	AA641937	ns18b10.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1183963 5', mRNA sequence.	Homo sapiens	35,586	27-OCT-1997
		GB_PR3:AC003074	143029	AC003074	Human PAC clone DJ0596O09 from 7p15, complete sequence.	Homo sapiens	31,917	6-Nov-97
rxa02106	1179	GB_BA1:SC1A6	37620	AL023496	Streptomyces coelicolor cosmid 1A6.	Streptomyces coelicolor	35,818	13-Jan-99
		GB_PR4:AC005553	179651	AC005553	Homo sapiens chromosome 17, clone hRPK.112_J_9, complete sequence.	Homo sapiens	34,274	31-DEC-1998
		GB_EST3:R49746	397	R49746	yg71g10.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:38768 5' similar to gb:V00567 BETA-2-MICROGLOBULIN PRECURSOR (HUMAN);, mRNA sequence.	Homo sapiens	41,162	18-MAY-1995
rxa02111	1407	GB_BA1:SC6G10	36734	AL049497	Streptomyces coelicolor cosmid 6G10.	Streptomyces coelicolor	50,791	24-MAR-1999
		GB_BA1:U00010	41171	U00010	Mycobacterium leprae cosmid B1170.	Mycobacterium leprae	37,563	01-MAR-1994
		GB_BA1:MTCY336	32437	Z95586	Mycobacterium tuberculosis H37Rv complete genome; segment 70/162.	Mycobacterium tuberculosis	39,504	24-Jun-99
rxa02112	960	GB_HTG3:AC010579	157658	AC010579	Drosophila melanogaster chromosome 3 clone BACR09D08 (D1101) RPCI-98 09.D.8 map 96F-96F strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 121 unordered pieces.	Drosophila melanogaster	37,909	24-Sep-99
		GB_GSS3:B09839	1191	B09839	T12A12-Sp6 TAMU Arabidopsis thaliana genomic clone T12A12, genomic survey sequence.	Arabidopsis thaliana	37,843	14-MAY-1997
		GB_HTG3:AC010579	157658	AC010579	Drosophila melanogaster chromosome 3 clone BACR09D08 (D1101) RPCI-98 09.D.8 map 96F-96F strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 121 unordered pieces.	Drosophila melanogaster	37,909	24-Sep-99
rxa02134	1044	GB_BA1:SCSECYDNA	6154	X83011	S.coelicolor secY locus DNA.	Streptomyces coelicolor	36,533	02-MAR-1998
		GB_EST32:AI731596	568	AI731596	BNLGHi10185 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (AC004005) putative ribosomal protein L7 [Arabidopsis thaliana], mRNA sequence.	Gossypium hirsutum	33,451	11-Jun-99
		GB_BA1:SCSECYDNA	6154	X83011	S.coelicolor secY locus DNA.	Streptomyces coelicolor	36,756	02-MAR-1998

TABLE 4: ALIGNMENT RESULTS

rxa02135	1197	GB_PR3:HS525L6	168111	AL023807	Human DNA sequence from clone RP3-525L6 on chromosome 6p22.3-23 Contains CA repeat, STSs, GSSs and a CpG Island, complete sequence.	Homo sapiens	34,365	23-Nov-99
		GB_PL2:ATF21P8	85785	AL022347	Arabidopsis thaliana DNA chromosome 4, BAC clone F21P8 (ESSA project).	Arabidopsis thaliana	34,325	9-Jun-99
		GB_PL2:U89959	106973	U89959	Arabidopsis thaliana BAC T7I23, complete sequence.	Arabidopsis thaliana	33,874	26-Jun-98
rxa02136	645	GB_PL2:ATAC005819	57752	AC005819	Arabidopsis thaliana chromosome II BAC T3A4 genomic sequence, complete sequence.	Arabidopsis thaliana	34,123	3-Nov-98
		GB_PL2:F15K9	71097	AC005278	Arabidopsis thaliana chromosome 1 BAC F15K9 sequence, complete sequence.	Arabidopsis thaliana	31,260	7-Nov-98
		GB_PL2:U89959	106973	U89959	Arabidopsis thaliana BAC T7I23, complete sequence.	Arabidopsis thaliana	34,281	26-Jun-98
rxa02139	1962	GB_BA1:MTCY190	34150	Z70283	Mycobacterium tuberculosis H37Rv complete genome; segment 98/162.	Mycobacterium tuberculosis	62,904	17-Jun-98
		GB_BA1:MSGB1554CS	36548	L78814	Mycobacterium leprae cosmid B1554 DNA sequence.	Mycobacterium leprae	36,648	15-Jun-96
		GB_BA1:MSGB1551CS	36548	L78813	Mycobacterium leprae cosmid B1551 DNA sequence.	Mycobacterium leprae	36,648	15-Jun-96
rxa02153	903	GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	99,104	1-Jul-98
		GB_BA1:AF005242	1044	AF005242	Corynebacterium glutamicum N-acetylglutamate-5-semialdehyde dehydrogenase (argC) gene, complete cds.	Corynebacterium glutamicum	99,224	2-Jul-97
		GB_BA1:CGARGCJBD	4355	X86157	C.glutamicum argC, argJ, argB, argD, and argF genes.	Corynebacterium glutamicum	100,000	25-Jul-96
rxa02154	414	GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	98,551	1-Jul-98
		GB_BA1:AF005242	1044	AF005242	Corynebacterium glutamicum N-acetylglutamate-5-semialdehyde dehydrogenase (argC) gene, complete cds.	Corynebacterium glutamicum	98,477	2-Jul-97
		GB_BA1:CGARGCJBD	4355	X86157	C.glutamicum argC, argJ, argB, argD, and argF genes.	Corynebacterium glutamicum	100,000	25-Jul-96
rxa02155	1287	GB_BA1:CGARGCJBD	4355	X86157	C.glutamicum argC, argJ, argB, argD, and argF genes.	Corynebacterium glutamicum	99,767	25-Jul-96
		GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	99,378	1-Jul-98
		GB_BA1:MSGB1133CS	42106	L78811	Mycobacterium leprae cosmid B1133 DNA sequence.	Mycobacterium leprae	55,504	15-Jun-96

TABLE 4: ALIGNMENT RESULTS

rxa02156	1074	GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	100,000	1-Jul-98
		GB_BA1:CGARGCJBD	4355	X86157	C.glutamicum argC, argJ, argB, argD, and argF genes.	Corynebacterium glutamicum	100,000	25-Jul-96
		GB_BA2:AE001816	10007	AE001816	Thermotoga maritima section 128 of 136 of the complete genome.	Thermotoga maritima	50,238	2-Jun-99
rxa02157	1296	GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	99,612	1-Jul-98
		GB_BA1:CGARGCJBD	4355	X86157	C.glutamicum argC, argJ, argB, argD, and argF genes.	Corynebacterium glutamicum	99,612	25-Jul-96
		GB_BA1:MTCY06H11	38000	Z85982	Mycobacterium tuberculosis H37Rv complete genome; segment 73/162.	Mycobacterium tuberculosis	57,278	17-Jun-98
rxa02158	1080	GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	100,000	1-Jul-98
		GB_BA2:AF031518	2045	AF031518	Corynebacterium glutamicum ornithine carbamolytransferase (argF) gene, complete cds.	Corynebacterium glutamicum	99,898	5-Jan-99
		GB_BA1:CGARGCJBD	4355	X86157	C.glutamicum argC, argJ, argB, argD, and argF genes.	Corynebacterium glutamicum	100,000	25-Jul-96
rxa02159	636	GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	99,843	1-Jul-98
		GB_BA2:AF031518	2045	AF031518	Corynebacterium glutamicum ornithine carbamolytransferase (argF) gene, complete cds.	Corynebacterium glutamicum	88,679	5-Jan-99
		GB_BA2:AF041436	516	AF041436	Corynebacterium glutamicum arginine repressor (argR) gene, complete cds.	Corynebacterium glutamicum	100,000	5-Jan-99
rxa02160	1326	GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	99,774	1-Jul-98
		GB_BA2:AF030520	1206	AF030520	Corynebacterium glutamicum argininosuccinate synthetase (argG) gene, complete cds.	Corynebacterium glutamicum	99,834	19-Nov-97
		GB_BA1:SCARGGH	1909	Z49111	S.clavuligerus argG gene and argH gene (partial).	Streptomyces clavuligerus	65,913	22-Apr-96

TABLE 4: ALIGNMENT RESULTS

rxa02162	1554	GB_BA2:AF049897	9196	AF049897	Corynebacterium glutamicum N-acetylglutamylphosphate reductase (argC), ornithine acetyltransferase (argJ), N-acetylglutamate kinase (argB), acetylornithine transaminase (argD), ornithine carbamoyltransferase (argF), arginine repressor (argR), argininosuccinate synthase (argG), and argininosuccinate lyase (argH) genes, complete cds.	Corynebacterium glutamicum	88,524	1-Jul-98
		GB_BA2:AF048764	1437	AF048764	Corynebacterium glutamicum argininosuccinate lyase (argH) gene,complete cds.	Corynebacterium glutamicum	87,561	1-Jul-98
		GB_BA1:MTCY06H11	38000	Z85982	Mycobacterium tuberculosis H37Rv complete genome; segment 73/162.	Mycobacterium tuberculosis	64,732	17-Jun-98
rxa02176	1251	GB_BA1:MTCY31	37630	Z73101	Mycobacterium tuberculosis H37Rv complete genome; segment 41/162.	Mycobacterium tuberculosis	36,998	17-Jun-98
		GB_BA1:CGGLTG	3013	X66112	C.glutamicum glt gene for citrate synthase and ORF.	Corynebacterium glutamicum	39,910	17-Feb-95
		GB_PL2:PGU65399	2700	U65399	Basidiomycete CECT 20197 phenoloxidase (pox1) gene, complete cds.	basidiomycete CECT 20197	38,474	19-Jul-97
rxa02189	861	GB_PR3:AC002468	115888	AC002468	Human Chromosome 15q26.1 PAC clone pDJ417d7, complete sequence.	Homo sapiens	35,941	16-Sep-98
		GB_BA1:MSGB1970CS	39399	L78815	Mycobacterium leprae cosmid B1970 DNA sequence.	Mycobacterium leprae	40,286	15-Jun-96
		GB_PR3:AC002468	115888	AC002468	Human Chromosome 15q26.1 PAC clone pDJ417d7, complete sequence.	Homo sapiens	33,689	16-Sep-98
rxa02193	1701	GB_BA1:BRLASPA	1987	D25316	Brevibacterium flavum aspA gene for aspartase, complete cds.	Corynebacterium glutamicum	99,353	6-Feb-99
		GB_PAT:E04307	1581	E04307	DNA encoding Brevibacterium flavum aspartase.	Corynebacterium glutamicum	99,367	29-Sep-97
		GB_BA1:ECOUW93	338534	U14003	Escherichia coli K-12 chromosomal region from 92.8 to 00.1 minutes.	Escherichia coli	37,651	17-Apr-96
rxa02194	966	GB_BA2:AF050166	840	AF050166	Corynebacterium glutamicum ATP phosphoribosyltransferase (hisG) gene, complete cds.	Corynebacterium glutamicum	98,214	5-Jan-99
		GB_BA1:BRLASPA	1987	D25316	Brevibacterium flavum aspA gene for aspartase, complete cds.	Corynebacterium glutamicum	93,805	6-Feb-99
		GB_PAT:E08649	188	E08649	DNA encoding part of aspartase from coryneform bacteria.	Corynebacterium glutamicum	100,000	29-Sep-97
rxa02195	393	GB_BA2:AF086704	264	AF086704	Corynebacterium glutamicum phosphoribosyl-ATP-pyrophosphohydrolase (hisE) gene, complete cds.	Corynebacterium glutamicum	100,000	8-Feb-99
		GB_BA1:EAY17145	6019	Y17145	Eubacterium acidaminophilum grdR, grdI, grdH genes and partial ldc, grdT genes.	Eubacterium acidaminophilum	39,075	5-Aug-98
		GB_STS:G01195	332	G01195	fruit fly STS Dm1930 clone DS06959 T7.	Drosophila melanogaster	35,542	28-Feb-95
rxa02197	551	GB_BA1:MTCY261	27322	Z97559	Mycobacterium tuberculosis H37Rv complete genome; segment 95/162.	Mycobacterium tuberculosis	33,938	17-Jun-98
		GB_BA1:MLCB2533	40245	AL035310	Mycobacterium leprae cosmid B2533.	Mycobacterium leprae	65,517	27-Aug-99
		GB_BA1:U00017	42157	U00017	Mycobacterium leprae cosmid B2126.	Mycobacterium leprae	36,770	01-MAR-1994
rxa02198	2599	GB_BA1:U00017	42157	U00017	Mycobacterium leprae cosmid B2126.	Mycobacterium leprae	38,674	01-MAR-1994
		GB_BA1:MLCB2533	40245	AL035310	Mycobacterium leprae cosmid B2533.	Mycobacterium leprae	65,465	27-Aug-99
		GB_BA1:MTCY261	27322	Z97559	Mycobacterium tuberculosis H37Rv complete genome; segment 95/162.	Mycobacterium tuberculosis	37,577	17-Jun-98
rxa02208	1025	GB_BA1:U00017	42157	U00017	Mycobacterium leprae cosmid B2126.	Mycobacterium leprae	59,823	01-MAR-1994
		GB_BA1:AP000063	185300	AP000063	Aeropyrum pernix genomic DNA, section 6/7.	Aeropyrum pernix	39,442	22-Jun-99
		GB_PR4:AC006236	127593	AC006236	Homo sapiens chromosome 17, clone hCIT.162_E_12, complete sequence.	Homo sapiens	37,191	29-DEC-1998
rxa02229	948	GB_BA1:MSGY154	40221	AD000002	Mycobacterium tuberculosis sequence from clone y154.	Mycobacterium tuberculosis	53,541	03-DEC-1996
		GB_BA1:MTCY154	13935	Z98209	Mycobacterium tuberculosis H37Rv complete genome; segment 121/162.	Mycobacterium tuberculosis	40,407	17-Jun-98
		GB_BA1:U00019	36033	U00019	Mycobacterium leprae cosmid B2235.	Mycobacterium leprae	40,541	01-MAR-1994
rxa02234	3462	GB_BA1:MSGB937CS	38914	L78820	Mycobacterium leprae cosmid B937 DNA sequence.	Mycobacterium leprae	66,027	15-Jun-96
		GB_BA1:MTCY2B12	20431	Z81011	Mycobacterium tuberculosis H37Rv complete genome; segment 61/162.	Mycobacterium tuberculosis	71,723	18-Jun-98
		GB_BA2:U01072	4393	U01072	Mycobacterium bovis BCG orotidine-5'-monophosphate decarboxylase (uraA) gene.	Mycobacterium bovis	67,101	22-DEC-1993

TABLE 4: ALIGNMENT RESULTS

rxa02235	727	GB_BA1:MSU91572	960	U91572	Mycobacterium smegmatis carbamoyl phosphate synthetase (pyrAB) gene, partial cds and orotidine 5'-monophosphate decarboxylase (pyrF) gene, complete cds.	Mycobacterium smegmatis	60,870	22-MAR-1997
		GB_HTG3:AC009364	192791	AC009364	Homo sapiens chromosome 7, *** SEQUENCING IN PROGRESS ***, 57 unordered pieces.	Homo sapiens	37,994	1-Sep-99
		GB_HTG3:AC009364	192791	AC009364	Homo sapiens chromosome 7, *** SEQUENCING IN PROGRESS ***, 57 unordered pieces.	Homo sapiens	37,994	1-Sep-99
rxa02237	693	GB_BA1:MTCY21B4	39150	Z80108	Mycobacterium tuberculosis H37Rv complete genome; segment 62/162.	Mycobacterium tuberculosis	55,844	23-Jun-98
		GB_BA2:AF077324	5228	AF077324	Rhodococcus equi strain 103 plasmid RE-VP1 fragment f.	Rhodococcus equi	41,185	5-Nov-98
		GB_EST22:AU017763	586	AU017763	AU017763 Mouse two-cell stage embryo cDNA Mus musculus cDNA clone J0744A04 3', mRNA sequence.	Mus musculus	38,616	19-OCT-1998
rxa02239	1389	GB_BA1:MTCY21B4	39150	Z80108	Mycobacterium tuberculosis H37Rv complete genome; segment 62/162.	Mycobacterium tuberculosis	56,282	23-Jun-98
		GB_HTG3:AC010745	193862	AC010745	Homo sapiens clone NH0549D18, *** SEQUENCING IN PROGRESS ***, 30 unordered pieces.	Homo sapiens	36,772	21-Sep-99
		GB_HTG3:AC010745	193862	AC010745	Homo sapiens clone NH0549D18, *** SEQUENCING IN PROGRESS ***, 30 unordered pieces.	Homo sapiens	36,772	21-Sep-99
rxa02240	1344	EM_PAT:E09855	1239	E09855	gDNA encoding S-adenosylmethionine synthetase.	Corynebacterium glutamicum	99,515	07-OCT-1997 (Rel. 52, Created)
		GB_PAT:A37831	5392	A37831	Sequence 1 from Patent WO9408014.	Streptomyces pristinaespiralis	63,568	05-MAR-1997
		GB_BA2:AF117274	2303	AF117274	Streptomyces spectabilis flavoprotein homolog Dfp (dfp) gene, partial cds; and S-adenosylmethionine synthetase (metK) gene, complete cds.	Streptomyces spectabilis	65,000	31-MAR-1999
rxa02246	1107	EM_BA1:AB003693	5589	AB003693	Corynebacterium ammoniagenes DNA for rib operon, complete cds.	Corynebacterium ammoniagenes	52,909	03-OCT-1997 (Rel. 52, Created)
		GB_PAT:E07957	5589	E07957	gDNA encoding at least guanosine triphosphate cyclohydrolase and riboflavin synthase.	Corynebacterium ammoniagenes	52,909	29-Sep-97
		GB_PAT:I32742	5589	I32742	Sequence 1 from patent US 5589355.	Unknown.	52,909	6-Feb-97
rxa02247	756	GB_PAT:I32743	2689	I32743	Sequence 2 from patent US 5589355.	Unknown.	57,937	6-Feb-97
		EM_BA1:AB003693	5589	AB003693	Corynebacterium ammoniagenes DNA for rib operon, complete cds.	Corynebacterium ammoniagenes	57,937	03-OCT-1997 (Rel. 52, Created)
		GB_PAT:I32742	5589	I32742	Sequence 1 from patent US 5589355.	Unknown.	57,937	6-Feb-97
rxa02248	1389	GB_PAT:I32742	5589	I32742	Sequence 1 from patent US 5589355.	Unknown.	61,843	6-Feb-97
		EM_BA1:AB003693	5589	AB003693	Corynebacterium ammoniagenes DNA for rib operon, complete cds.	Corynebacterium ammoniagenes	61,843	03-OCT-1997 (Rel. 52, Created)
		GB_PAT:E07957	5589	E07957	gDNA encoding at least guanosine triphosphate cyclohydrolase and riboflavin synthase.	Corynebacterium ammoniagenes	61,843	29-Sep-97
rxa02249	600	GB_PAT:E07957	5589	E07957	gDNA encoding at least guanosine triphosphate cyclohydrolase and riboflavin synthase.	Corynebacterium ammoniagenes	64,346	29-Sep-97
		GB_PAT:I32742	5589	I32742	Sequence 1 from patent US 5589355.	Unknown.	64,346	6-Feb-97
		GB_PAT:I32743	2689	I32743	Sequence 2 from patent US 5589355.	Unknown.	64,346	6-Feb-97

TABLE 4: ALIGNMENT RESULTS

rxa02250	643	GB_PAT:E07957	5589	E07957	gDNA encoding at least guanosine triphosphate cyclohydrolase and riboflavin synthase.	Corynebacterium ammoniagenes	56,318	29-Sep-97
		GB_PAT:I32742	5589	I32742	Sequence 1 from patent US 5589355.	Unknown.	56,318	6-Feb-97
		EM_BA1:AB003693	5589	AB003693	Corynebacterium ammoniagenes DNA for rib operon, complete cds.	Corynebacterium ammoniagenes	56,318	03-OCT-1997 (Rel. 52, Created)
rxa02262	1269	GB_BA1:CGL007732	4460	AJ007732	Corynebacterium glutamicum 3' ppc gene, secG gene, amt gene, ocd gene and 5' soxA gene.	Corynebacterium glutamicum	100,000	7-Jan-99
		GB_BA1:CGAMTGENE	2028	X93513	C.glutamicum amt gene.	Corynebacterium glutamicum	100,000	29-MAY-1996
		GB_VI:HEHCMVCG	229354	X17403	Human cytomegalovirus strain AD169 complete genome.	human herpesvirus 5	38,651	10-Feb-99
rxa02263	488	GB_BA1:CGL007732	4460	AJ007732	Corynebacterium glutamicum 3' ppc gene, secG gene, amt gene, ocd gene and 5' soxA gene.	Corynebacterium glutamicum	100,000	7-Jan-99
		GB_BA1:CGL007732	4460	AJ007732	Corynebacterium glutamicum 3' ppc gene, secG gene, amt gene, ocd gene and 5' soxA gene.	Corynebacterium glutamicum	37,526	7-Jan-99
rxa02272	1368	EM_PAT:E09373	1591	E09373	Creatinine deiminase gene.	Bacillus sp.	96,928	08-OCT-1997 (Rel. 52, Created)
		GB_BA1:D38505	1591	D38505	Bacillus sp. gene for creatinine deaminase, complete cds.	Bacillus sp.	96,781	7-Aug-98
		GB_HTG2:AC006595	146070	AC006595	Homo sapiens, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.	Homo sapiens	36,264	20-Feb-99
rxa02281	1545	GB_GSS12:AQ411010	551	AQ411010	HS_2257_B1_H02_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2257 Col=3 Row=P, genomic survey sequence.	Homo sapiens	36,197	17-MAR-1999
		GB_EST23:AI128623	363	AI128623	qa62c01.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:1691328 3', mRNA sequence.	Homo sapiens	37,017	05-OCT-1998
		GB_PL2:ATAC007019	102335	AC007019	Arabidopsis thaliana chromosome II BAC F7D8 genomic sequence, complete sequence.	Arabidopsis thaliana	33,988	16-MAR-1999
rxa02299	531	GB_BA2:AF116184	540	AF116184	Corynebacterium glutamicum L-aspartate-alpha-decarboxylase precursor (panD) gene, complete cds.	Corynebacterium glutamicum	100,000	02-MAY-1999
		GB_GSS9:AQ164310	507	AQ164310	HS_2171_A2_E01_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2171 Col=2 Row=I, genomic survey sequence.	Homo sapiens	37,278	16-OCT-1998
		GB_VI:MH68TKH	4557	X93468	Murine herpesvirus type 68 thymidine kinase and glycoprotein H genes.	murine herpesvirus 68	40,288	3-Sep-96
rxa02311	813	GB_HTG4:AC006091	176878	AC006091	Drosophila melanogaster chromosome 3 clone BACR48G05 (D475) RPCI-98 48.G.5 map 91F1-91F13 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.	Drosophila melanogaster	36,454	27-OCT-1999
		GB_HTG4:AC006091	176878	AC006091	Drosophila melanogaster chromosome 3 clone BACR48G05 (D475) RPCI-98 48.G.5 map 91F1-91F13 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.	Drosophila melanogaster	36,454	27-OCT-1999

TABLE 4: ALIGNMENT RESULTS

		GB_BA2:RRU65510	16259	U65510	Rhodospirillum rubrum CO-induced hydrogenase operon (cooM, cooK, cooL, cooX, cooU, cooH) genes, iron sulfur protein (cooF) gene, carbon monoxide dehydrogenase (cooS) gene, carbon monoxide dehydrogenase accessory proteins (cooC, cooT, cooJ) genes, putative transcriptional activator (cooA) gene, nicotinate-nucleotide pyrophosphorylase (nadC) gene, complete cds, L-aspartate oxidase (nadB) gene, and alkyl hydroperoxide reductase (ahpC) gene, partial cds.	Rhodospirillum rubrum	37,828	9-Apr-97
rxa02315	1752	GB_BA1:MSGY224	40051	AD000004	Mycobacterium tuberculosis sequence from clone y224.	Mycobacterium tuberculosis	49,418	03-DEC-1996
		GB_BA1:MTY25D10	40838	Z95558	Mycobacterium tuberculosis H37Rv complete genome; segment 28/162.	Mycobacterium tuberculosis	49,360	17-Jun-98
		GB_BA1:MSGY224	40051	AD000004	Mycobacterium tuberculosis sequence from clone y224.	Mycobacterium tuberculosis	38,150	03-DEC-1996
rxa02318	402	GB_HTG3:AC011348	111083	AC011348	Homo sapiens chromosome 5 clone CIT-HSPC_303E13, *** SEQUENCING IN PROGRESS ***, 3 ordered pieces.	Homo sapiens	35,821	06-OCT-1999
		GB_HTG3:AC011348	111083	AC011348	Homo sapiens chromosome 5 clone CIT-HSPC_303E13, *** SEQUENCING IN PROGRESS ***, 3 ordered pieces.	Homo sapiens	35,821	06-OCT-1999
		GB_HTG3:AC011412	89234	AC011412	Homo sapiens chromosome 5 clone CIT978SKB_81K21, *** SEQUENCING IN PROGRESS ***, 3 ordered pieces.	Homo sapiens	36,181	06-OCT-1999
rxa02319	1080	GB_BA1:MSGY224	40051	AD000004	Mycobacterium tuberculosis sequence from clone y224.	Mycobacterium tuberculosis	37,792	03-DEC-1996
		GB_BA1:MTY25D10	40838	Z95558	Mycobacterium tuberculosis H37Rv complete genome; segment 28/162.	Mycobacterium tuberculosis	37,792	17-Jun-98
		GB_EST23:AI117213	476	AI117213	ub83h02.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:1395123 5',mRNA sequence.	Mus musculus	35,084	2-Sep-98
rxa02345	1320	GB_BA1:BAPURKE	2582	X91189	B.ammoniagenes purK and purE genes.	Corynebacterium ammoniagenes	61,731	14-Jan-97
		GB_BA1:MTCY71	42729	Z92771	Mycobacterium tuberculosis H37Rv complete genome; segment 141/162.	Mycobacterium tuberculosis	39,624	10-Feb-99
		GB_BA1:MTCY71	42729	Z92771	Mycobacterium tuberculosis H37Rv complete genome; segment 141/162.	Mycobacterium tuberculosis	39,847	10-Feb-99
rxa02350	618	GB_BA1:BAPURKE	2582	X91189	B.ammoniagenes purK and purE genes.	Corynebacterium ammoniagenes	64,286	14-Jan-97
		GB_PL1:SC130KBXV	129528	X94335	S.cerevisiae 130kb DNA fragment from chromosome XV.	Saccharomyces cerevisiae	36,617	15-Jul-97
		GB_PL1:SCXVORFS	50984	X90518	S.cerevisiae DNA of 51 Kb from chromosome XV right arm.	Saccharomyces cerevisiae	36,617	1-Nov-95
rxa02373	1038	GB_PAT:E00311	1853	E00311	DNA coding of 2,5-diketogluconic acid reductase.	unidentified	56,123	29-Sep-97
		GB_PAT:I06030	1853	I06030	Sequence 4 from Patent EP 0305608.	Unknown.	56,220	02-DEC-1994
		GB_PAT:I00836	1853	I00836	Sequence 1 from Patent US 4758514.	Unknown.	56,220	21-MAY-1993
rxa02375	1350	GB_BA2:CGU31230	3005	U31230	Corynebacterium glutamicum Obg protein homolog gene, partial cds, gamma glutamyl kinase (proB) gene, complete cds, and (unkdh) gene, complete cds.	Corynebacterium glutamicum	99,332	2-Aug-96
		GB_HTG3:AC009946	169072	AC009946	Homo sapiens clone NH0012C17, *** SEQUENCING IN PROGRESS ***, 1 unordered pieces.	Homo sapiens	36,115	8-Sep-99
		GB_HTG3:AC009946	169072	AC009946	Homo sapiens clone NH0012C17, *** SEQUENCING IN PROGRESS ***, 1 unordered pieces.	Homo sapiens	36,115	8-Sep-99
rxa02380	777	GB_BA1:MTCY253	41230	Z81368	Mycobacterium tuberculosis H37Rv complete genome; segment 106/162.	Mycobacterium tuberculosis	38,088	17-Jun-98
		GB_HTG4:AC010658	120754	AC010658	Drosophila melanogaster chromosome 3L/75C1 clone RPCI98-3B20, *** SEQUENCING IN PROGRESS ***, 78 unordered pieces.	Drosophila melanogaster	35,817	16-OCT-1999
		GB_HTG4:AC010658	120754	AC010658	Drosophila melanogaster chromosome 3L/75C1 clone RPCI98-3B20, *** SEQUENCING IN PROGRESS ***, 78 unordered pieces.	Drosophila melanogaster	35,817	16-OCT-1999

TABLE 4: ALIGNMENT RESULTS

rxa02382	1419	GB_BA1:CGPROAGEN	1783	X82929	C.glutamicum proA gene.	Corynebacterium glutamicum	98,802	23-Jan-97
		GB_BA1:MTCY428	26914	Z81451	Mycobacterium tuberculosis H37Rv complete genome; segment 107/162.	Mycobacterium tuberculosis	38,054	17-Jun-98
		GB_BA2:CGU31230	3005	U31230	Corynebacterium glutamicum Obg protein homolog gene, partial cds, gamma glutamyl kinase (proB) gene, complete cds, and (unkdh) gene, complete cds.	Corynebacterium glutamicum	98,529	2-Aug-96
rxa02400	693	GB_BA1:CGACEA	2427	X75504	C.glutamicum aceA gene and thiX genes (partial).	Corynebacterium glutamicum	100,000	9-Sep-94
		GB_PAT:I86191	2135	I86191	Sequence 3 from patent US 5700661.	Unknown.	100,000	10-Jun-98
		GB_PAT:I13693	2135	I13693	Sequence 3 from patent US 5439822.	Unknown.	100,000	26-Sep-95
rxa02432	1098	GB_GSS15:AQ606842	574	AQ606842	HS_5404_B2_E07_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=980 Col=14 Row=J, genomic survey sequence.	Homo sapiens	39,716	10-Jun-99
		GB_EST1:T05804	406	T05804	EST03693 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBDG63 similar to EST containing Alu repeat, mRNA sequence.	Homo sapiens	37,915	30-Jun-93
		GB_PL1:AB006699	77363	AB006699	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MDJ22, complete sequence.	Arabidopsis thaliana	35,526	20-Nov-99
rxa02458	1413	GB_BA2:AF114233	1852	AF114233	Corynebacterium glutamicum 5-enolpyruvylshikimate 3-phosphate synthase (aroA) gene, complete cds.	Corynebacterium glutamicum	100,000	7-Feb-99
		GB_EST37:AW013061	578	AW013061	ODT-0033 Winter flounder ovary Pleuronectes americanus cDNA clone ODT-0033 5' similar to FRUCTOSE-BISPHOSPHATE ALDOLASE B (LIVER), mRNA sequence.	Pleuronectes americanus	39,175	10-Sep-99
		GB_GSS15:AQ650027	728	AQ650027	Sheared DNA-5L2.TF Sheared DNA Trypanosoma brucei genomic clone Sheared DNA-5L2, genomic survey sequence.	Trypanosoma brucei	39,281	22-Jun-99
rxa02469	1554	GB_BA1:MTCY359	36021	Z83859	Mycobacterium tuberculosis H37Rv complete genome; segment 84/162.	Mycobacterium tuberculosis	39,634	17-Jun-98
		GB_BA1:MLCB1788	39228	AL008609	Mycobacterium leprae cosmid B1788.	Mycobacterium leprae	59,343	27-Aug-99
		GB_BA1:SCAJ10601	4692	AJ010601	Streptomyces coelicolor A3(2) DNA for whiD and whiK loci.	Streptomyces coelicolor	48,899	17-Sep-98
rxa02497	1050	GB_BA2:CGU31224	422	U31224	Corynebacterium glutamicum (ppx) gene, partial cds.	Corynebacterium glutamicum	96,445	2-Aug-96
		GB_BA1:MTCY20G9	37218	Z77162	Mycobacterium tuberculosis H37Rv complete genome; segment 25/162.	Mycobacterium tuberculosis	59,429	17-Jun-98
		GB_BA1:SCE7	16911	AL049819	Streptomyces coelicolor cosmid E7.	Streptomyces coelicolor	39,510	10-MAY-1999
rxa02499	933	GB_BA2:CGU31225	1817	U31225	Corynebacterium glutamicum L-proline:NADP+ 5-oxidoreductase (proC) gene, complete cds.	Corynebacterium glutamicum	97,749	2-Aug-96
		GB_BA1:NG17PILA	1920	X13965	Neisseria gonorrhoeae pilA gene.	Neisseria gonorrhoeae	43,249	30-Sep-93
		GB_HTG2:AC007984	129715	AC007984	Drosophila melanogaster chromosome 3 clone BACR05C10 (D781) RPCI-98 05.C.10 map 97D-97E strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 87 unordered pieces.	Drosophila melanogaster	33,406	2-Aug-99
rxa02501	1188	GB_BA1:MTCY20G9	37218	Z77162	Mycobacterium tuberculosis H37Rv complete genome; segment 25/162.	Mycobacterium tuberculosis	39,357	17-Jun-98
		GB_BA1:U00018	42991	U00018	Mycobacterium leprae cosmid B2168.	Mycobacterium leprae	51,768	01-MAR-1994
		GB_VI:HE1CG	152261	X14112	Herpes simplex virus (HSV) type 1 complete genome.	human herpesvirus 1	39,378	17-Apr-97
rxa02503	522	GB_PR3:AC005328	35414	AC005328	Homo sapiens chromosome 19, cosmid R26660, complete sequence.	Homo sapiens	39,922	28-Jul-98
		GB_PR3:AC005545	43514	AC005545	Homo sapiens chromosome 19, cosmid R26634, complete sequence.	Homo sapiens	39,922	3-Sep-98
		GB_PR3:AC005328	35414	AC005328	Homo sapiens chromosome 19, cosmid R26660, complete sequence.	Homo sapiens	34,911	28-Jul-98
rxa02504	681	GB_BA1:MTCY20G9	37218	Z77162	Mycobacterium tuberculosis H37Rv complete genome; segment 25/162.	Mycobacterium tuberculosis	54,940	17-Jun-98
		GB_PR3:AC005328	35414	AC005328	Homo sapiens chromosome 19, cosmid R26660, complete sequence.	Homo sapiens	41,265	28-Jul-98
		GB_PR3:AC005545	43514	AC005545	Homo sapiens chromosome 19, cosmid R26634, complete sequence.	Homo sapiens	41,265	3-Sep-98
rxa02516	1386	GB_BA1:MLCL536	36224	Z99125	Mycobacterium leprae cosmid L536.	Mycobacterium leprae	37,723	04-DEC-1998

TABLE 4: ALIGNMENT RESULTS

		GB_BA1:U00013	35881	U00013	Mycobacterium leprae cosmid B1496.	Mycobacterium leprae	37,723	01-MAR-1994
		GB_BA1:MTV007	32806	AL021184	Mycobacterium tuberculosis H37Rv complete genome; segment 64/162.	Mycobacterium tuberculosis	61,335	17-Jun-98
rxa02517	570	GB_BA1:MLCL536	36224	Z99125	Mycobacterium leprae cosmid L536.	Mycobacterium leprae	37,018	04-DEC-1998
		GB_BA1:U00013	35881	U00013	Mycobacterium leprae cosmid B1496.	Mycobacterium leprae	37,018	01-MAR-1994
		GB_BA1:SCC22	22115	AL096839	Streptomyces coelicolor cosmid C22.	Streptomyces coelicolor	37,071	12-Jul-99
rxa02532	1170	GB_OV:AF137219	831	AF137219	Amia calva mixed lineage leukemia-like protein (Mll) gene, partial cds.	Amia calva	36,853	7-Sep-99
		GB_EST30:AI645057	301	AI645057	vs52a10.y1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone IMAGE:1149882 5', mRNA sequence.	Mus musculus	41,860	29-Apr-99
		GB_EST20:AA822595	429	AA822595	vs52a10.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone IMAGE:1149882 5', mRNA sequence.	Mus musculus	42,353	17-Feb-98
rxa02536	879	GB_HTG2:AF130866	118874	AF130866	Homo sapiens chromosome 8 clone PAC 172N13 map 8q24, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	40,754	21-MAR-1999
		GB_HTG2:AF130866	118874	AF130866	Homo sapiens chromosome 8 clone PAC 172N13 map 8q24, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	40,754	21-MAR-1999
		GB_PL1:ATT12J5	84499	AL035522	Arabidopsis thaliana DNA chromosome 4, BAC clone T12J5 (ESSAII project).	Arabidopsis thaliana	35,063	24-Feb-99
rxa02550	1434	GB_BA1:MTCY279	9150	Z97991	Mycobacterium tuberculosis H37Rv complete genome; segment 17/162.	Mycobacterium tuberculosis	37,773	17-Jun-98
		GB_BA1:MSGB1970CS	39399	L78815	Mycobacterium leprae cosmid B1970 DNA sequence.	Mycobacterium leprae	39,024	15-Jun-96
		GB_BA2:SC2H4	25970	AL031514	Streptomyces coelicolor cosmid 2H4.	Streptomyces coelicolor A3(2)	37,906	19-OCT-1999
rxa02559	1026	GB_BA1:MTV004	69350	AL009198	Mycobacterium tuberculosis H37Rv complete genome; segment 144/162.	Mycobacterium tuberculosis	47,358	18-Jun-98
		GB_PAT:I28684	5100	I28684	Sequence 1 from patent US 5573915.	Unknown.	39,138	6-Feb-97
		GB_BA1:MTU27357	5100	U27357	Mycobacterium tuberculosis cyclopropane mycolic acid synthase (cma1) gene, complete cds.	Mycobacterium tuberculosis	39,138	26-Sep-95
rxa02622	1683	GB_BA2:AE001780	11997	AE001780	Thermotoga maritima section 92 of 136 of the complete genome.	Thermotoga maritima	44,914	2-Jun-99
		GB_OV:AF064564	49254	AF064564	Fugu rubripes neurofibromatosis type 1 (NF1), A-kinase anchor protein (AKAP84), BAW protein (BAW), and WSB1 protein (WSB1) genes, complete cds.	Fugu rubripes	39,732	17-Aug-99
		GB_OV:AF064564	49254	AF064564	Fugu rubripes neurofibromatosis type 1 (NF1), A-kinase anchor protein (AKAP84), BAW protein (BAW), and WSB1 protein (WSB1) genes, complete cds.	Fugu rubripes	36,703	17-Aug-99
rxa02623	714	GB_GSS5:AQ818728	444	AQ818728	HS_5268_A1_G09_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=844 Col=17 Row=M, genomic survey sequence.	Homo sapiens	38,801	26-Aug-99
		GB_HTG5:AC011083	198586	AC011083	Homo sapiens chromosome 9 clone RP11-111M7 map 9, WORKING DRAFT SEQUENCE, 51 unordered pieces.	Homo sapiens	35,714	19-Nov-99
		GB_GSS6:AQ826948	544	AQ826948	HS_5014_A2_C12_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=590 Col=24 Row=E, genomic survey sequence.	Homo sapiens	39,146	27-Aug-99
rxa02629	708	GB_VI:BRSMGP	462	M86652	Bovine respiratory syncytial virus membrane glycoprotein mRNA, complete cds.	Bovine respiratory syncytial virus	37,013	28-Apr-93
		GB_VI:BRSMGP	462	M86652	Bovine respiratory syncytial virus membrane glycoprotein mRNA, complete cds.	Bovine respiratory syncytial virus	37,013	28-Apr-93

TABLE 4: ALIGNMENT RESULTS

rxa02645	1953	GB_PAT:A45577	1925	A45577	Sequence 1 from Patent WO9519442.	Corynebacterium glutamicum	39,130	07-MAR-1997
		GB_PAT:A45581	1925	A45581	Sequence 5 from Patent WO9519442.	Corynebacterium glutamicum	39,130	07-MAR-1997
		GB_BA1:CORILVA	1925	L01508	Corynebacterium glutamicum threonine dehydratase (ilvA) gene, complete cds.	Corynebacterium glutamicum	39,130	26-Apr-93
rxa02646	1392	GB_BA1:CORILVA	1925	L01508	Corynebacterium glutamicum threonine dehydratase (ilvA) gene, complete cds.	Corynebacterium glutamicum	99,138	26-Apr-93
		GB_PAT:A45585	1925	A45585	Sequence 9 from Patent WO9519442.	Corynebacterium glutamicum	99,066	07-MAR-1997
		GB_PAT:A45583	1925	A45583	Sequence 7 from Patent WO9519442.	Corynebacterium glutamicum	99,066	07-MAR-1997
rxa02648	1326	GB_OV:ICTCNC	2049	M83111	Ictalurus punctatus cyclic nucleotide-gated channel RNA sequence.	Ictalurus punctatus	38,402	24-MAY-1993
		GB_EST11:AA265464	345	AA265464	mx91c06.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:693706 5', mRNA sequence.	Mus musculus	38,655	20-MAR-1997
		GB_GSS8:AQ006950	480	AQ006950	CIT-HSP-2294E14.TR CIT-HSP Homo sapiens genomic clone 2294E14, genomic survey sequence.	Homo sapiens	36,074	27-Jun-98
rxa02653								
rxa02687	1068	GB_BA1:CORPHEA	1088	M13774	C.glutamicum pheA gene encoding prephenate dehydratase, complete cds.	Corynebacterium glutamicum	99,715	26-Apr-93
		GB_PAT:E04483	948	E04483	DNA encoding prephenate dehydratase.	Corynebacterium glutamicum	98,523	29-Sep-97
		GB_PAT:E06110	948	E06110	DNA encoding prephenate dehydratase.	Corynebacterium glutamicum	98,523	29-Sep-97
rxa02717	1005	GB_PL1:HVCH4H	59748	Y14573	Hordeum vulgare DNA for chromosome 4H.	Hordeum vulgare	36,593	25-MAR-1999
		GB_PR2:HS310H5	29718	Z69705	Human DNA sequence from cosmid 310H5 from a contig from the tip of the short arm of chromosome 16, spanning 2Mb of 16p13.3. Contains EST and CpG island.	Homo sapiens	36,089	22-Nov-99
		GB_PR3:AC004754	39188	AC004754	Homo sapiens chromosome 16, cosmid clone RT286 (LANL), complete sequence.	Homo sapiens	36,089	28-MAY-1998
rxa02754	1461	GB_HTG2:AC008223	130212	AC008223	Drosophila melanogaster chromosome 3 clone BACR16I18 (D815) RPCI-98 16.I.18 map 95A-95A strain y; cn bw sp, *** SEQUENCING IN PROGRESS***, 101 unordered pieces.	Drosophila melanogaster	32,757	2-Aug-99
		GB_HTG2:AC008223	130212	AC008223	Drosophila melanogaster chromosome 3 clone BACR16I18 (D815) RPCI-98 16.I.18 map 95A-95A strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 101 unordered pieces.	Drosophila melanogaster	32,757	2-Aug-99
		GB_BA1:MTCY71	42729	Z92771	Mycobacterium tuberculosis H37Rv complete genome; segment 141/162.	Mycobacterium tuberculosis	37,838	10-Feb-99
rxa02758	1422	GB_HTG5:AC011678	171967	AC011678	Homo sapiens clone 14_B_7, *** SEQUENCING IN PROGRESS ***, 20 unordered pieces.	Homo sapiens	35,331	5-Nov-99
		GB_HTG5:AC011678	171967	AC011678	Homo sapiens clone 14_B_7, *** SEQUENCING IN PROGRESS ***, 20 unordered pieces.	Homo sapiens	33,807	5-Nov-99
		GB_BA2:AF064070	23183	AF064070	Burkholderia pseudomallei putative dihydroorotase (pyrC) gene, partial cds; putative 1-acyl-sn-glycerol-3-phosphate acyltransferase (plsC), putative diadenosine tetraphosphatase (apaH), complete cds; type II O-antigen biosynthesis gene cluster, complete sequence; putative undecaprenyl phosphate N-acetylglucosaminyltransferase, and putative UDP-glucose 4-epimerase genes, complete cds; and putative galactosyl transferase gene, partial cds.	Burkholderia pseudomallei	36,929	20-Jan-99

TABLE 4: ALIGNMENT RESULTS

rxa02771	678	GB_BA2:AF038651	4077	AF038651	Corynebacterium glutamicum dipeptide-binding protein (dciAE) gene, partial cds; adenine phosphoribosyltransferase (apt) and GTP pyrophosphokinase (rel) genes, complete cds; and unknown gene.	Corynebacterium glutamicum	99,852	14-Sep-98
		GB_IN1:CELT19B4	37121	U80438	Caenorhabditis elegans cosmid T19B4.	Caenorhabditis elegans	43,836	04-DEC-1996
		GB_EST36:AV193572	360	AV193572	AV193572 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo Caenorhabditis elegans cDNA clone yk618h8 5', mRNA sequence.	Caenorhabditis elegans	48,588	22-Jul-99
rxa02772	1158	GB_BA2:AF038651	4077	AF038651	Corynebacterium glutamicum dipeptide-binding protein (dciAE) gene, partial cds; adenine phosphoribosyltransferase (apt) and GTP pyrophosphokinase (rel) genes, complete cds; and unknown gene.	Corynebacterium glutamicum	99,914	14-Sep-98
		GB_BA1:MTCY227	35946	Z77724	Mycobacterium tuberculosis H37Rv complete genome; segment 114/162.	Mycobacterium tuberculosis	38,339	17-Jun-98
		GB_BA1:U00011	40429	U00011	Mycobacterium leprae cosmid B1177.	Mycobacterium leprae	38,996	01-MAR-1994
rxa02790	1266	GB_BA1:MTCY159	33818	Z83863	Mycobacterium tuberculosis H37Rv complete genome; segment 111/162.	Mycobacterium tuberculosis	37,640	17-Jun-98
		GB_PR4:AC006581	172931	AC006581	Homo sapiens 12p21 BAC RPCI11-259O18 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	Homo sapiens	37,906	3-Jun-99
		GB_PR4:AC006581	172931	AC006581	Homo sapiens 12p21 BAC RPCI11-259O18 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	Homo sapiens	35,280	3-Jun-99
rxa02791	951	GB_BA1:MTCY159	33818	Z83863	Mycobacterium tuberculosis H37Rv complete genome; segment 111/162.	Mycobacterium tuberculosis	39,765	17-Jun-98
		GB_OV:CHKCEK2	3694	M35195	Chicken tyrosine kinase (cek2) mRNA, complete cds.	Gallus gallus	38,937	28-Apr-93
		GB_BA1:MSASDASK	5037	Z17372	M.smegmatis asd, ask-alpha, and ask-beta genes.	Mycobacterium smegmatis	38,495	9-Aug-94
rxa02802	1194	GB_EST24:AI223401	169	AI223401	qg48g01.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1838448 3' similar to WP:C25D7.8 CE08394 ;, mRNA sequence.	Homo sapiens	40,828	27-OCT-1998
		GB_EST24:AI223401	169	AI223401	qg48g01.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1838448 3' similar to WP:C25D7.8 CE08394 ;, mRNA sequence.	Homo sapiens	40,828	27-OCT-1998
rxa02814	494	GB_BA1:MTCY7D11	22070	Z95120	Mycobacterium tuberculosis H37Rv complete genome; segment 138/162.	Mycobacterium tuberculosis	58,418	17-Jun-98
		GB_BA1:MTCY7D11	22070	Z95120	Mycobacterium tuberculosis H37Rv complete genome; segment 138/162.	Mycobacterium tuberculosis	40,496	17-Jun-98
		GB_PR1:HSAJ2962	778	AJ002962	Homo sapiens mRNA for hB-FABP.	Homo sapiens	39,826	8-Jan-98
rxa02843	608	GB_BA1:CGAJ4934	1160	AJ004934	Corynebacterium glutamicum dapD gene, complete CDS.	Corynebacterium glutamicum	100,000	17-Jun-98
		GB_BA1:MTCI364	29540	Z93777	Mycobacterium tuberculosis H37Rv complete genome; segment 52/162.	Mycobacterium tuberculosis	37,710	17-Jun-98
		GB_BA1:MLU15180	38675	U15180	Mycobacterium leprae cosmid B1756.	Mycobacterium leprae	39,626	09-MAR-1995
rxs03205	963	GB_BA1:BLSIGBGN	2906	Z49824	B.lactofermentum orf1 gene and sigB gene.	Corynebacterium glutamicum	98,854	25-Apr-96
		GB_EST21:AA980237	377	AA980237	ua32a12.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1348414 5' similar to TR:Q61025 Q61025 HYPOTHETICAL 15.2 KD PROTEIN. ;, mRNA sequence.	Mus musculus	41,489	27-MAY-1998
		GB_EST23:AI158316	371	AI158316	ud27c05.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:1447112 5', mRNA sequence.	Mus musculus	38,005	30-Sep-98
rxs03223	1237	GB_IN1:LMFL2743	38368	AL031910	Leishmania major Friedlin chromosome 4 cosmid L2743.	Leishmania major	39,869	15-DEC-1999
		GB_PR3:HSDJ61B2	119666	AL096710	Human DNA sequence from clone RP1-61B2 on chromosome 6p11.2-12.3 Contains isoforms 1 and 3 of BPAG1 (bullous pemphigoid antigen 1 (230/240kD), an exon of a gene similar to murine MACF cytoskeletal protein, STSs and GSSs, complete sequence.	Homo sapiens	34,930	17-DEC-1999

TABLE 4: ALIGNMENT RESULTS

GB_PR3:HSDJ61B2	119666	AL096710	Human DNA sequence from clone RP1-61B2 on chromosome 6p11.2-12.3 Contains isoforms 1 and 3 of BPAG1 (bullous pemphigoid antigen 1 (230/240kD), an exon of a gene similar to murine MACF cytoskeletal protein, STSs and GSSs, complete sequence.	Homo sapiens	34.634	17-DEC-1999